NOAA Earth System
Research Laboratory

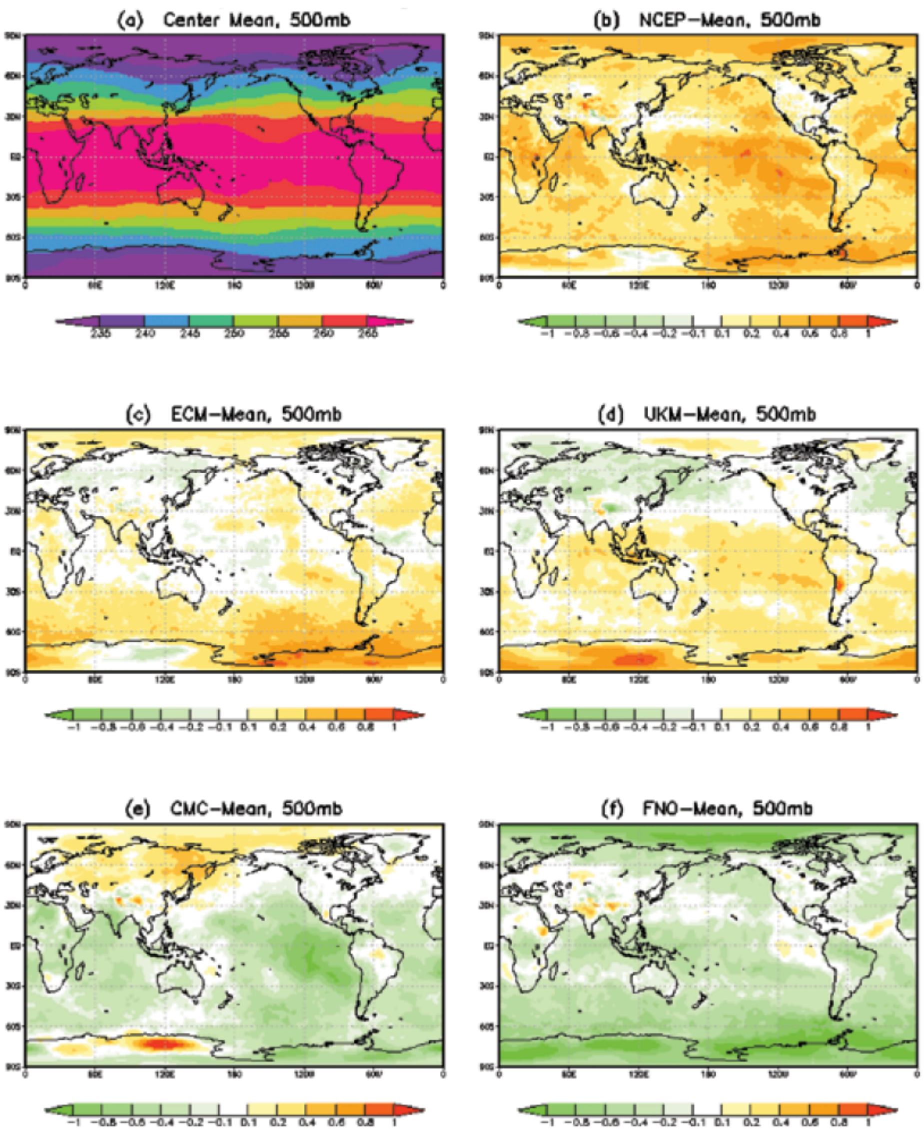
# Examining characteristics of analysis and short-range forecast errors using TIGGE

Tom Hamill
*NOAA Earth System Research Lab*
*Physical Sciences Division, Boulder CO*
(303) 497-3060; tom.hamill@noaa.gov



1 Dec 2011 presentation to NCEP/EMC/GMB. Acknowledgments: Daryl Kleist, Jeff Whitaker

# Original motivation




There are systematic differences between analyses from different centers, indicating that analyses are likely biased.

Supposition from WMO/JWGV:
since short-range forecasts might inherit analysis bias, and forecast bias may not be fully removed in next analysis, it's difficult to compare two centers' forecasts against one single center's analysis.

What then are the guiding principles, if any, for use of analysis data in verification?

from Wei et al. 2010

# When I started playing with the data, I started asking different questions.

- What can we understand about the characteristics of analysis errors by comparing analyses from different centers (supplied via THORPEX's new TIGGE database) ?  [important for ensemble initialization]
- Can we learn anything about potential forecast model and assimilation system biases from examining analyses and short-range forecasts using TIGGE's multi-center data?
- Assuming differences between centers' analyses are representative of analysis error, how well do various ensemble perturbation methodologies do in mimicking the properties of analysis errors?

# Some related literature


- Wei, M., Toth, Z., Zhu, Y., 2010: Analysis differences and error variance estimates from multi-center analysis data. *Australian Meteorological and Oceanographic Journal*, **59**, 25-34. pdf
- Langland, R.H., Maue, R.N. and Bishop, C.H. 2008. Uncertainty in atmospheric temperature analysis. *Tellus, **60A***, 598-603.
- Park, Y.-Y., Buizza, R. and Leutbecher, M. 2008. TIGGE: Preliminary results on comparing and combining ensembles. *QJRMS,* **134**, 2029-50.
- Bougeault, P., et al., 2010:The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull Amer. Meteor. Soc.*, **91**, 1059-1072. [link]

# Data and analysis methods

- I downloaded 00Z analyses, perturbed initial conditions, and deterministic forecasts at 2.5 degree resolution from ECMWF's TIGGE portal.
- Period is 1 Oct 2010 to 30 Sep 2011, 1 full year.
- Data from NCEP, ECMWF, UKMO, CMC, CMA (China).
- Variables: 2-meter temp, 10-m wind components, sea level pressure; temperature, wind components, and geopotential height at 50, 200, 250, 300, 500, 700, 850, and 925 hPa.
- Data to be shown
  - Time average of daily spread (sample standard deviation) of analyses about their daily mean.
  - Time series of analyses for select variables, at selected grid points, raw data and with +/- 15 day smooth.
  - 24-h forecast time tendencies, and how they differ from analysis tendencies, to get a sense as to whether the forecast model contributes a systematic bias that may be hard to remove in the analyses.
  - The power spectra of analysis differences, and how differences in perturbed initial conditions resemble these spectra.

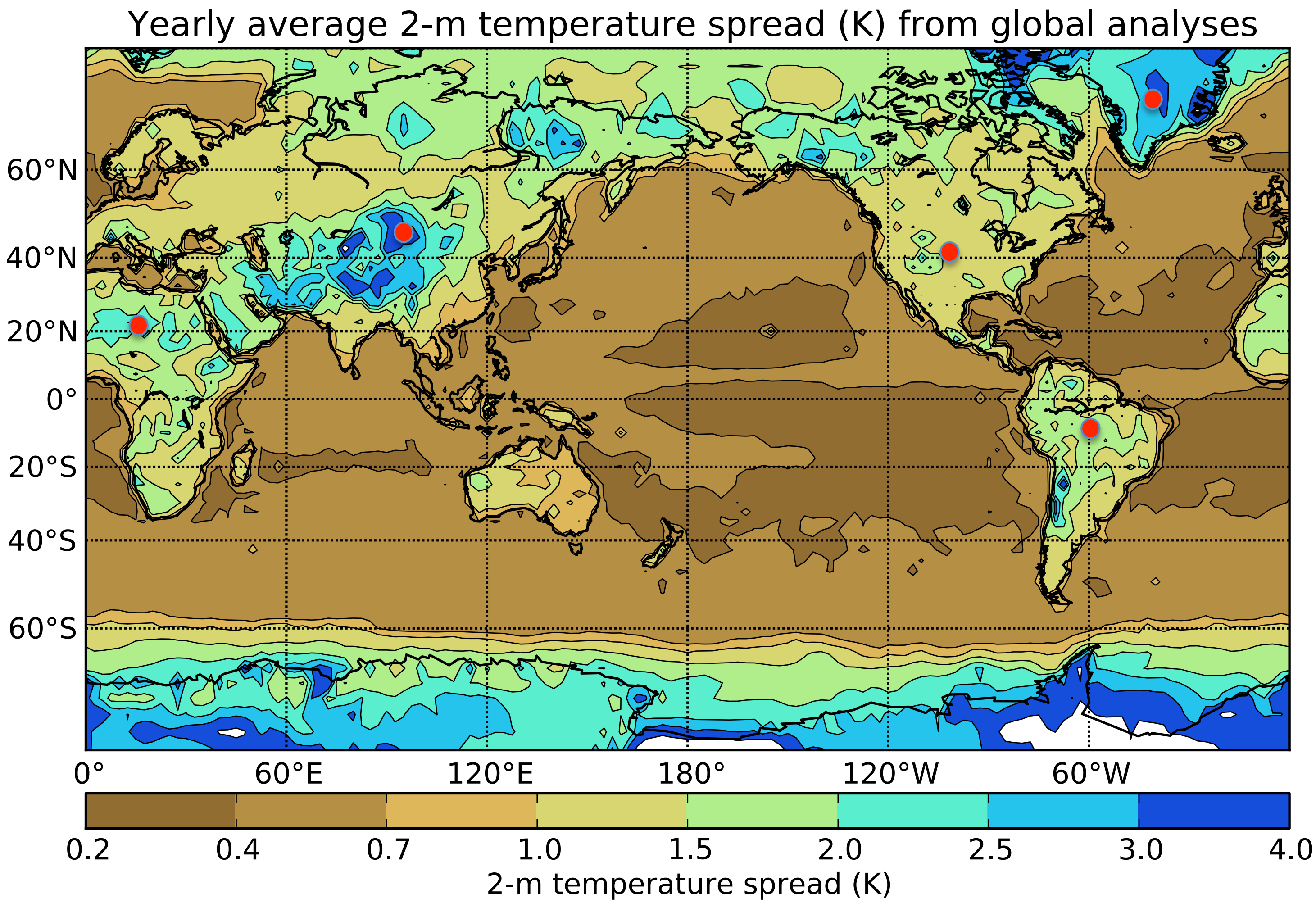
# Assumptions of this study

- Individual centers' analyses will have biases (almost certainly true, as will be shown).
- Biases between various analyses are uncorrelated, so the mean of multiple centers is less biased (unclear how much this is true ).
- Hence, the difference of a given center from the multi-center mean may indicate something about the bias in that analysis system.

Let’s look at 2-m temperatures…

# Analysis spread, 2-meter temperature

- red dots are locations of time series in subsequent plots

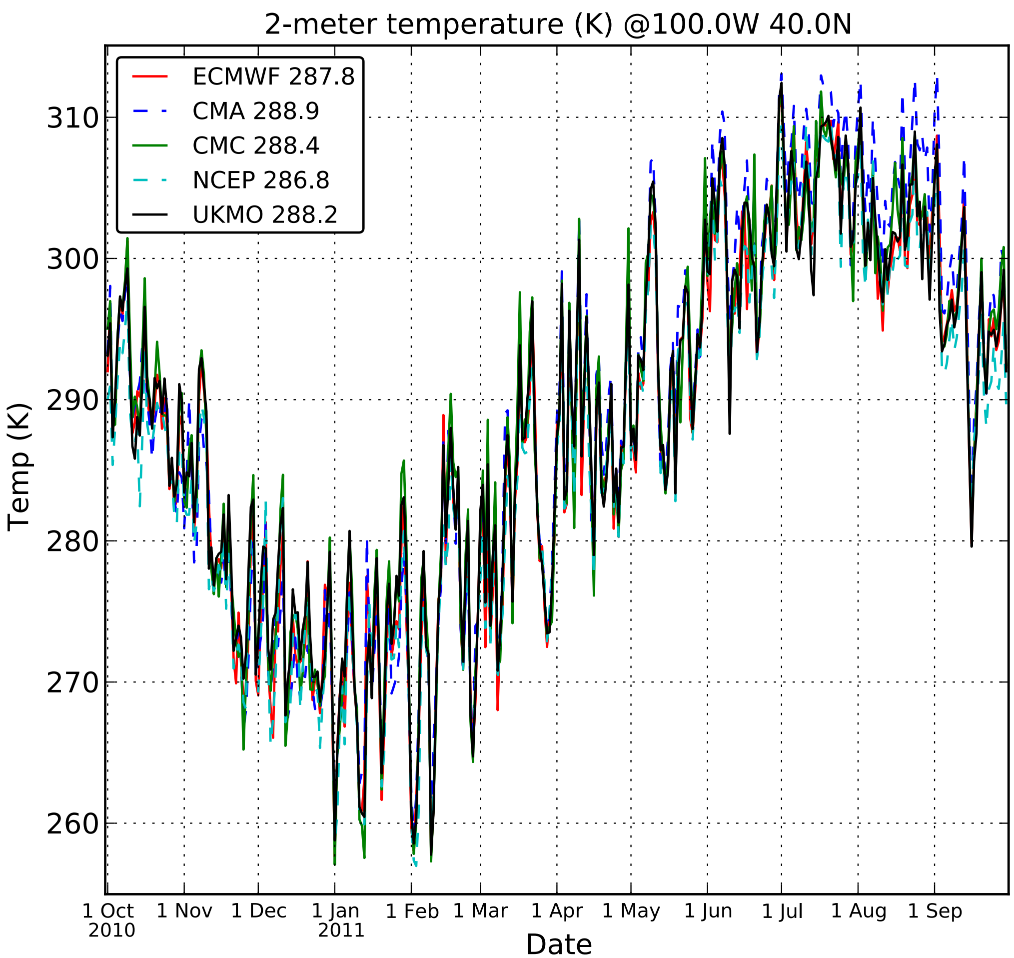



Here, I calculated the spread (sample std. dev.) on each day among the 5 analyses. The plot is the yearly average of these daily spreads.

> 1K spread over continents is typical, even in data dense regions.

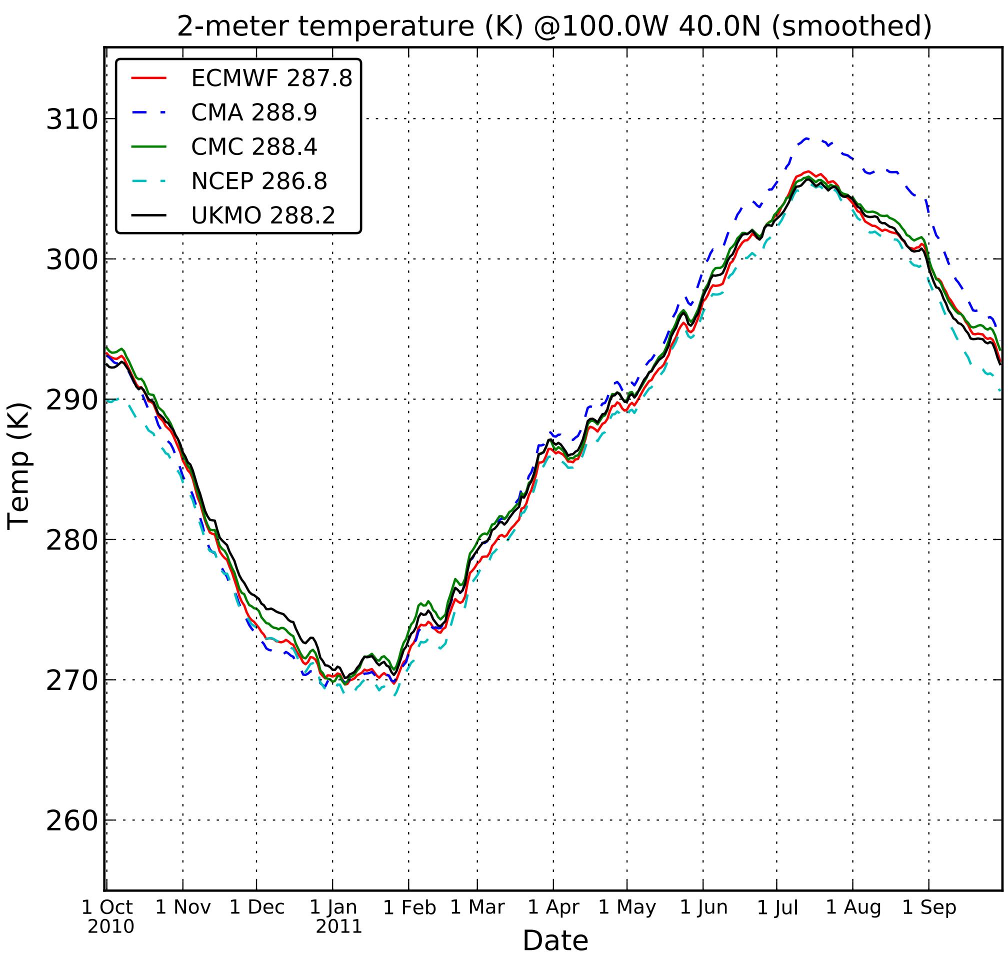
see also Fig 4. of Wei et al. 2010 ref.

# Time series of analyses, central US




Looking at this plot, the various line colors overlap quite a bit, suggesting that the differences have a substantial random component. However, looking at the yearly averaged temperature (listed in the box in the upper left), notice for example that NCEP's analysis is > 2 K colder than CMA's, on average.

# Time series of analyses, central US (smoothed)

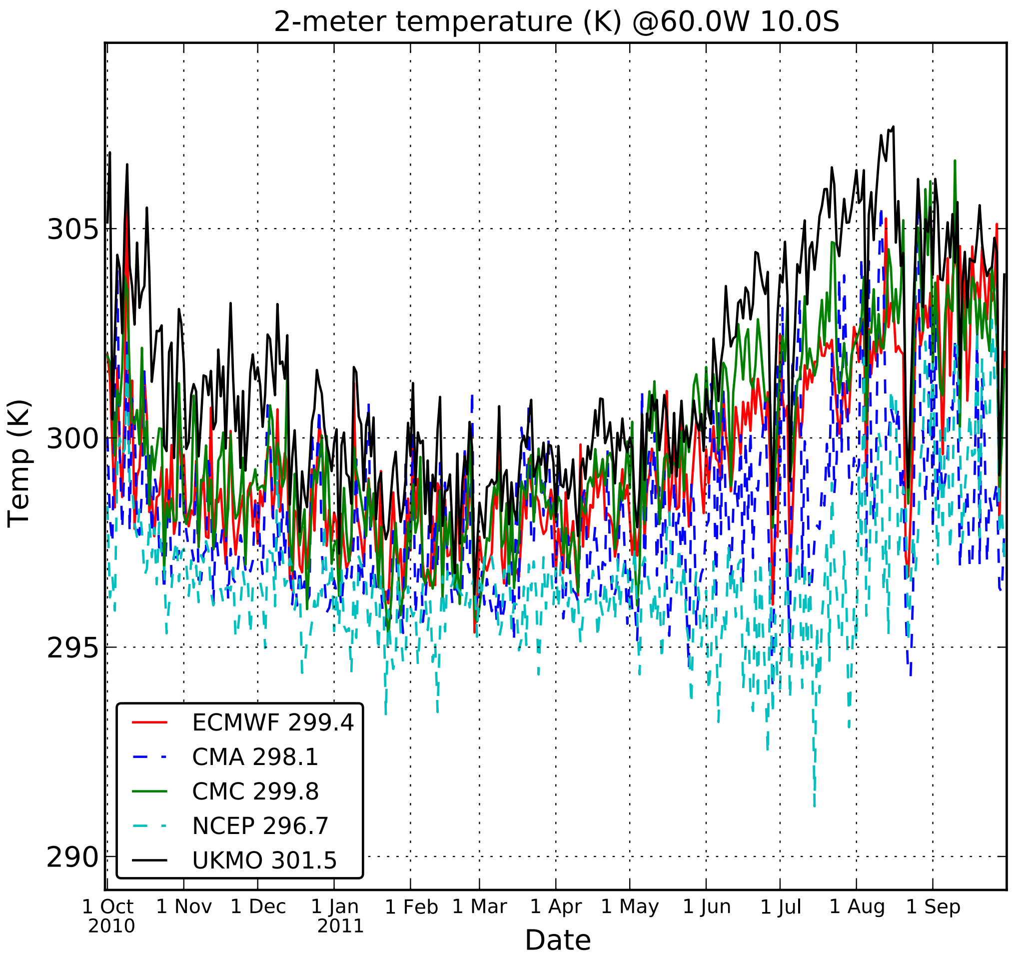



Here smoothed using average of +/- 15 days.

Warmer CMA analyses in last 4 months stands out.

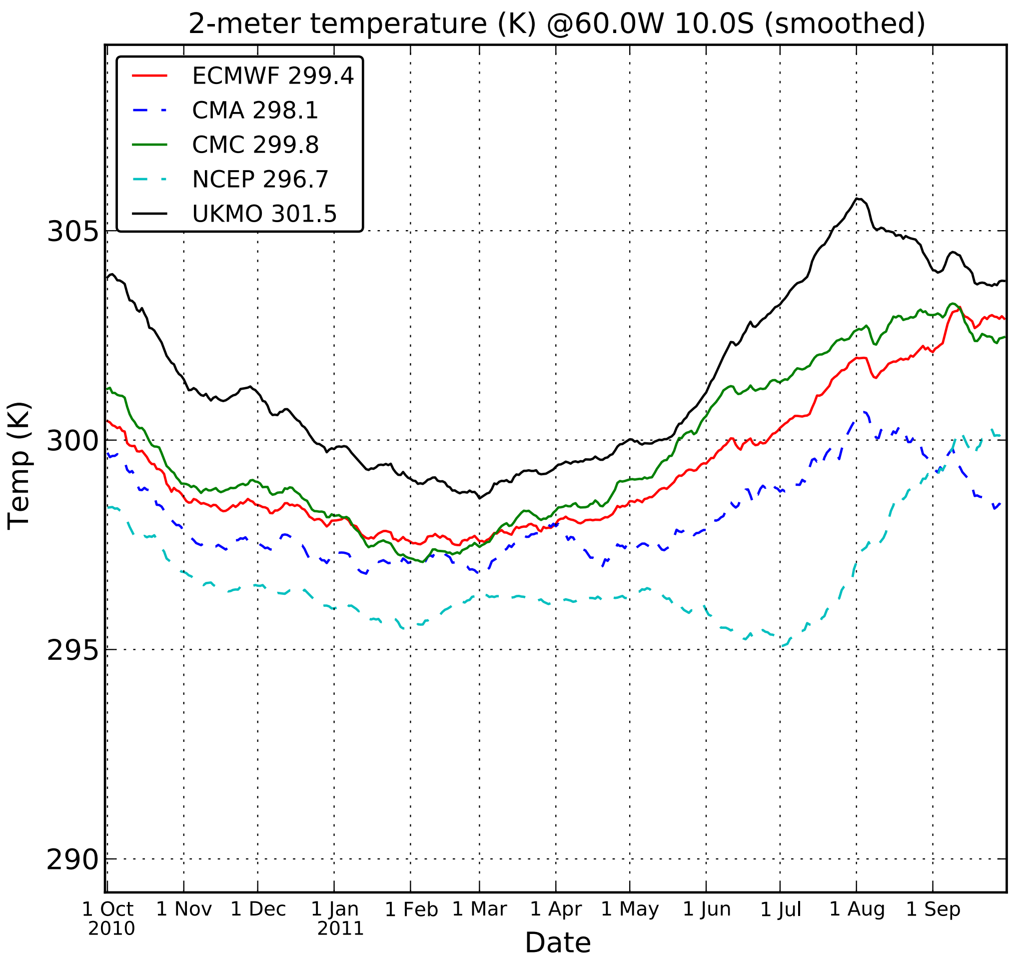
Even in a relatively data-rich region, there are apparent biases in analyses that may complicate their use for forecast verification.

# Time series of analyses, Amazon basin, South America



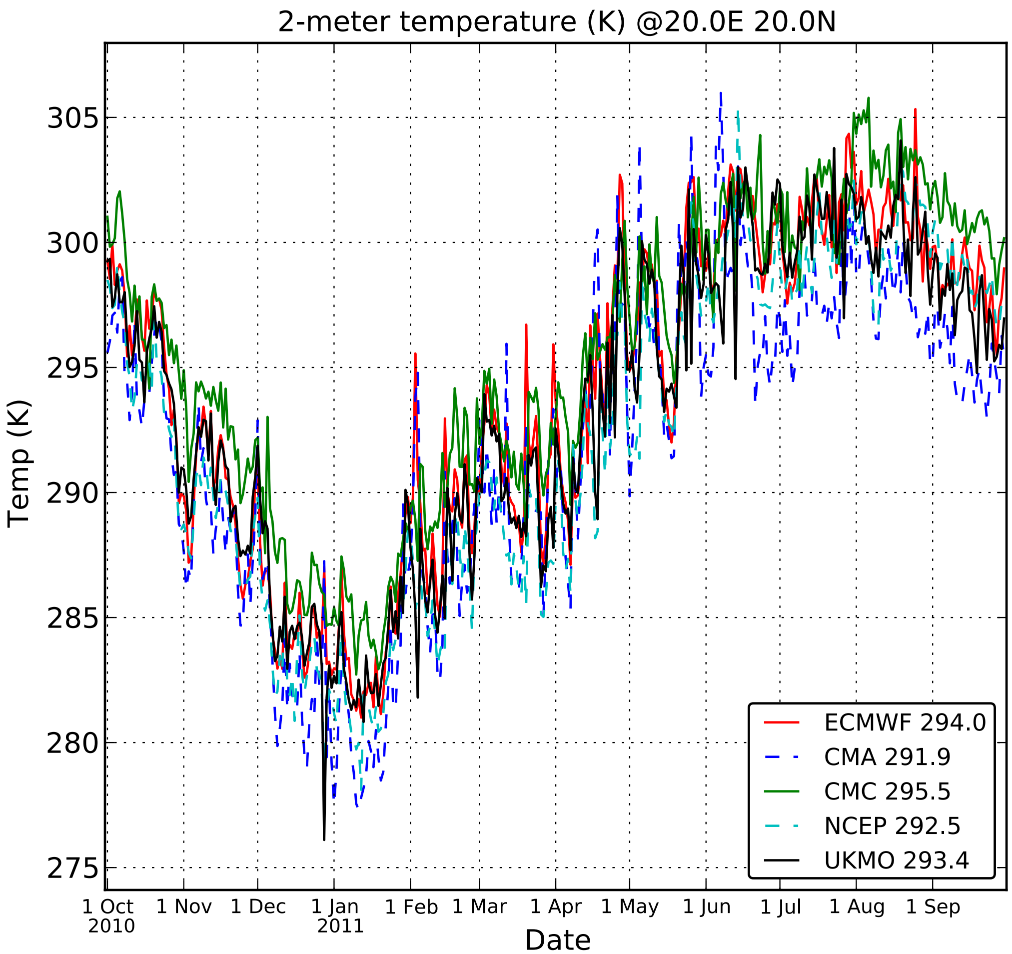

At this point in the Amazon basin, the analyses are quite different from each other, even in **yearly means**, which range from 296.7K to 301.5K. NCEP is rather consistently the coldest, UKMO the warmest. This suggests that a primary difference between analyses, especially in relatively data-sparse regions, may be a systematic bias in each analysis system.

# Time series of analyses, Amazon basin, South America, smoothed



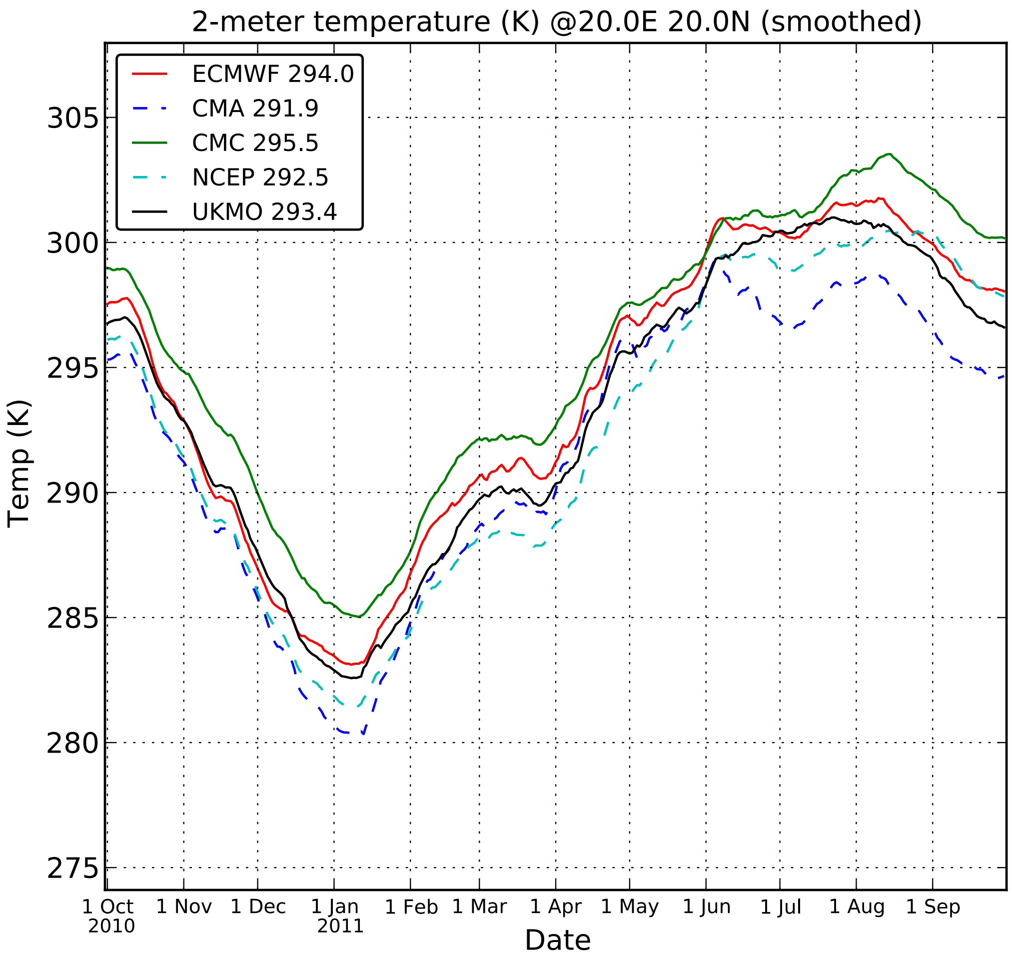

Both the warmer UKMO and cooler NCEP analyses stand out here.

# Time series of analyses, southern Saharan Desert, Africa



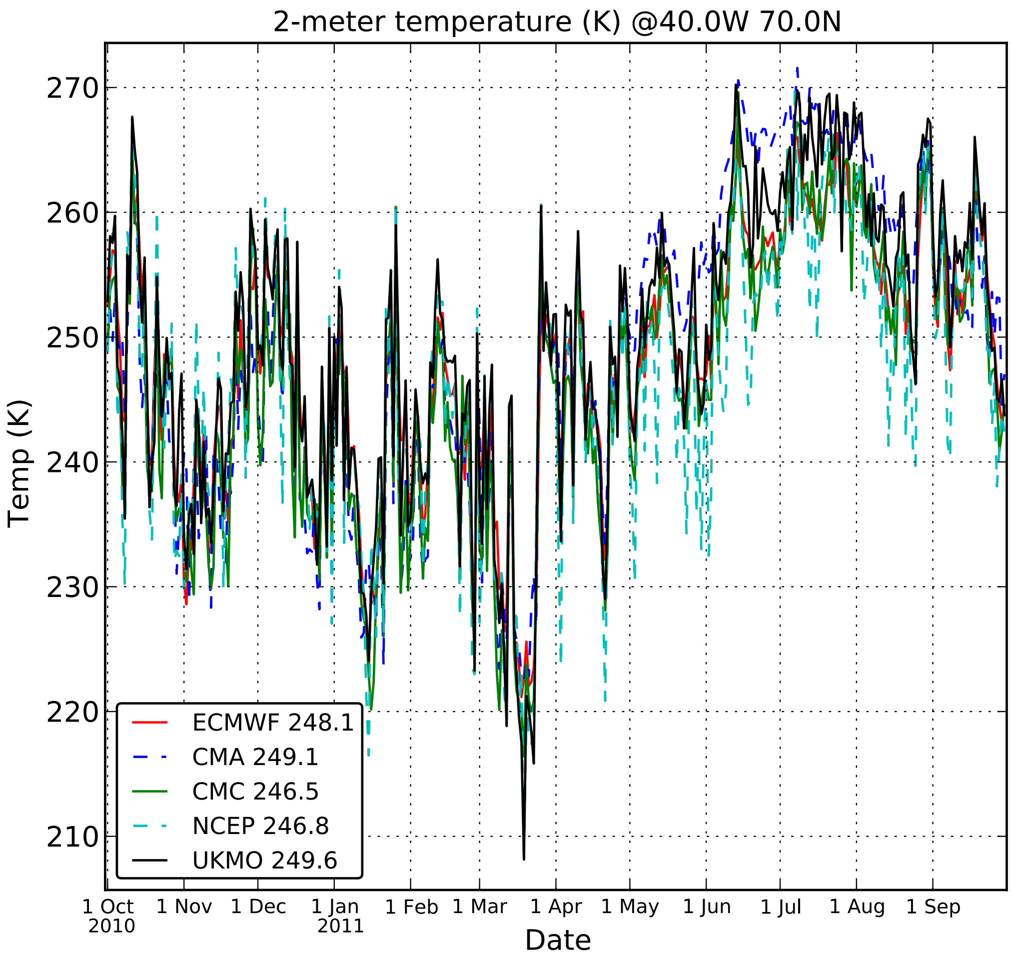

The relative order of which analysis is colder, which is warmer is different than in the Amazon, but there still appears that a substantial portion of the variability may be explained by systematic differences in the mean.

# Time series of analyses, southern Saharan Desert, Africa



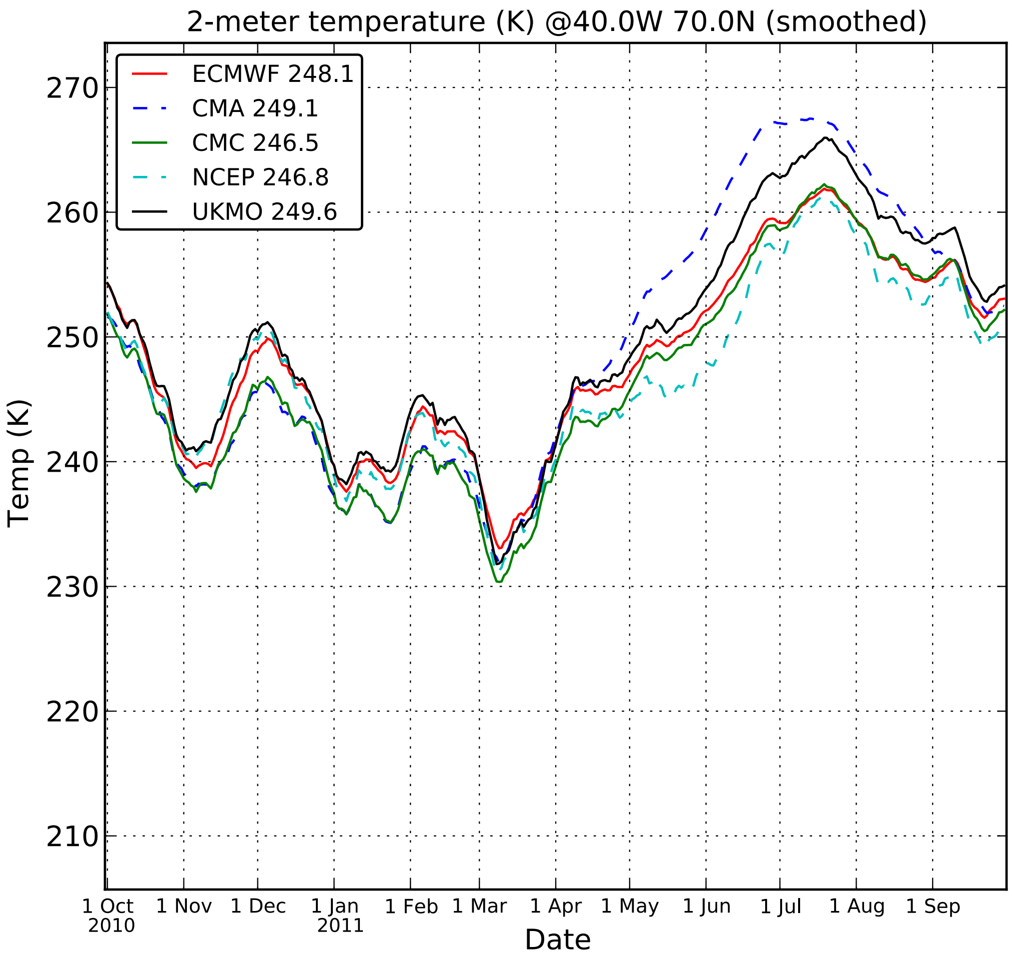

The rather consistent differences in the analyses is highlighted with the 30-day mean; CMC is rather consistently the warmest, CMA the coolest.

# Time series of analyses, central Greenland

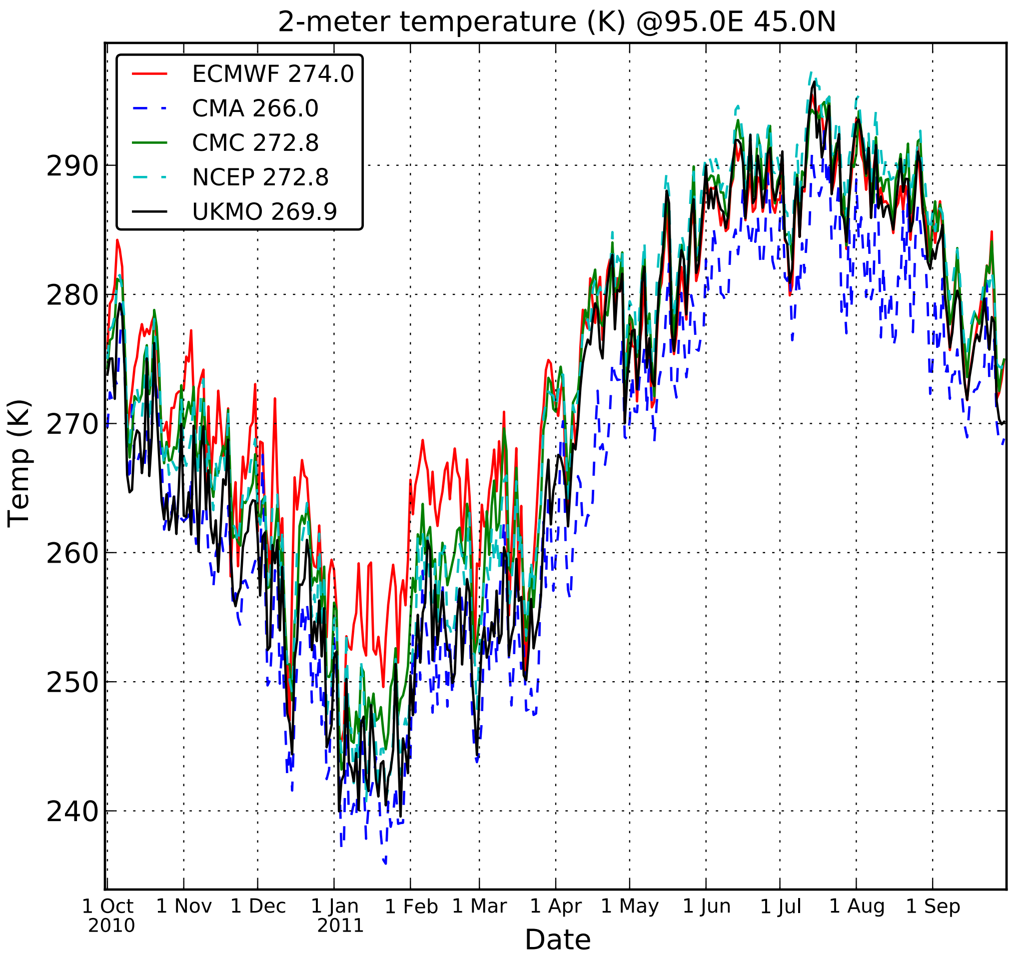



The analyses agree, Greenland was not the place you wanted to vacation on St. Patrick's Day in 2011 (March 17).

# Time series of analyses, central Greenland (smoothed)

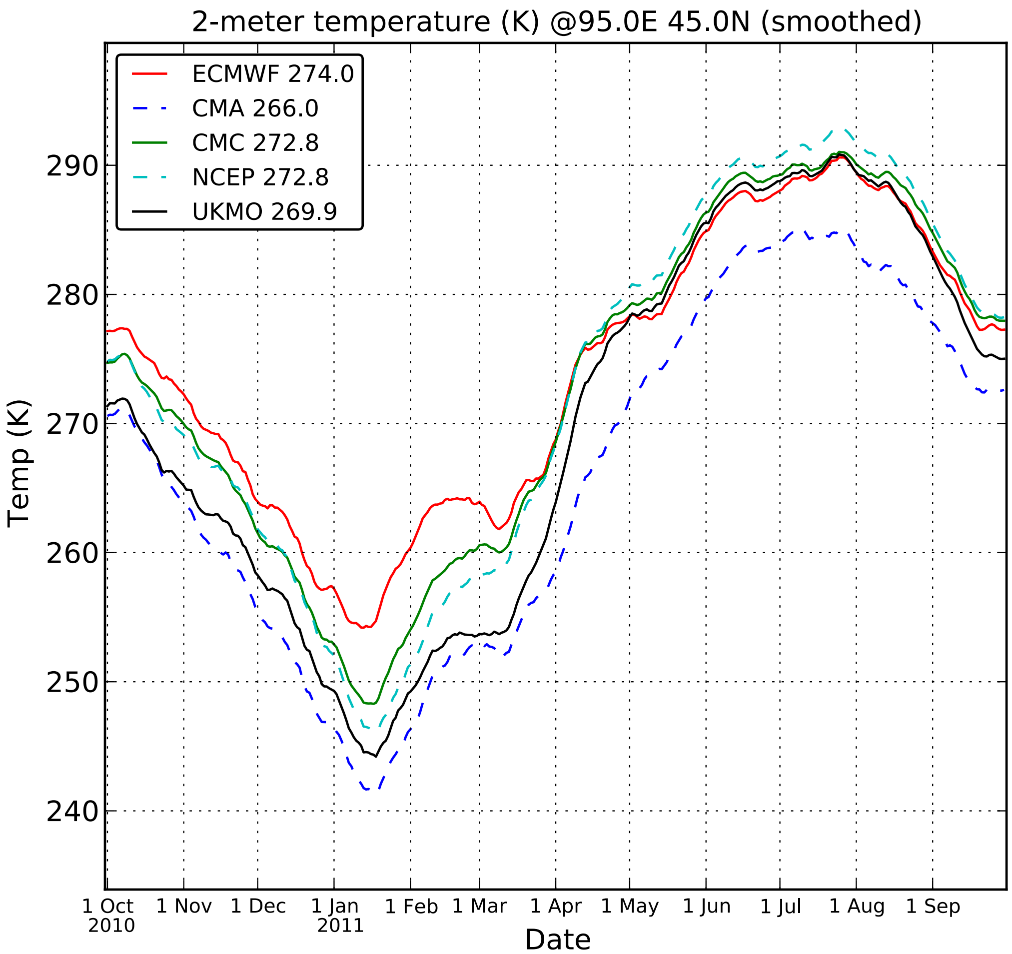


# Time series of analyses, Mongolia




Does ECMWF have the best analyses? They certainly didn't agree with any other center on the temperature in Mongolia last winter. The Chinese were the iconoclasts during the Mongolian summer.

# Time series of analyses, Mongolia



# Sources of bias in analyses

- Model bias inherited from prior first-guess forecast.
- Different methods of processing satellite data, i.e., different bias correction algorithms, different cloud clearing algorithms.
- Poor models of error covariances.
- Different data assimilated, perhaps due to different QC methods, data cutoffs.
- Etc.

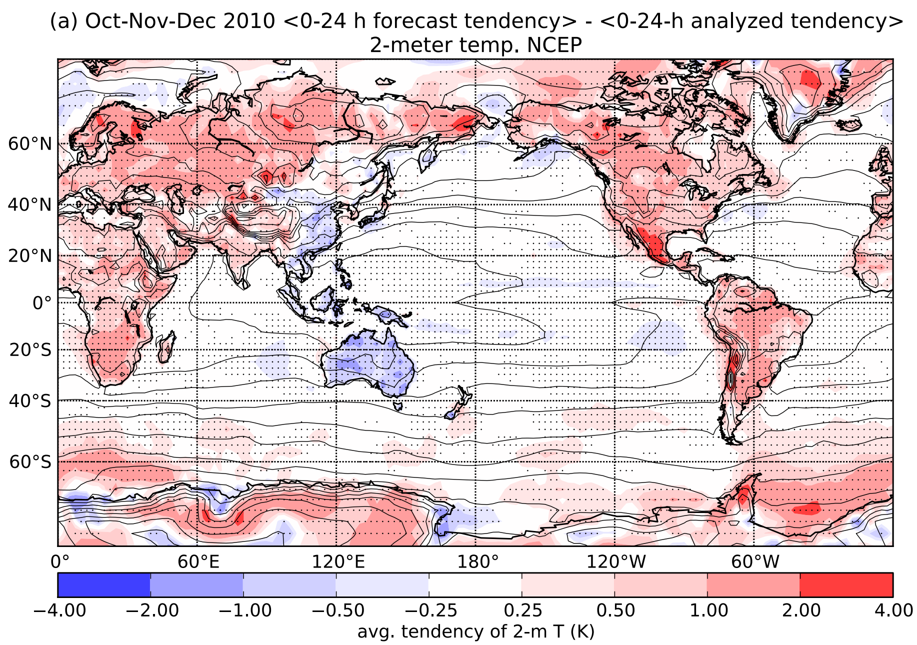
# Using 24-h time tendencies to diagnose possible model/DA system bias

- Neglecting the change of seasons, one would expect a long time series of $<F_{t+24} - A_t>$ would be unbiased, i.e., no systematic tendencies in 1-day forecasts. $< \cdot >$ is time average.
- Can account for change of seasons by calculating $<F_{t+24} - A_t>$ but then subtracting $<A_{t+24} - A_t>$.
- Subsequent plots show this.

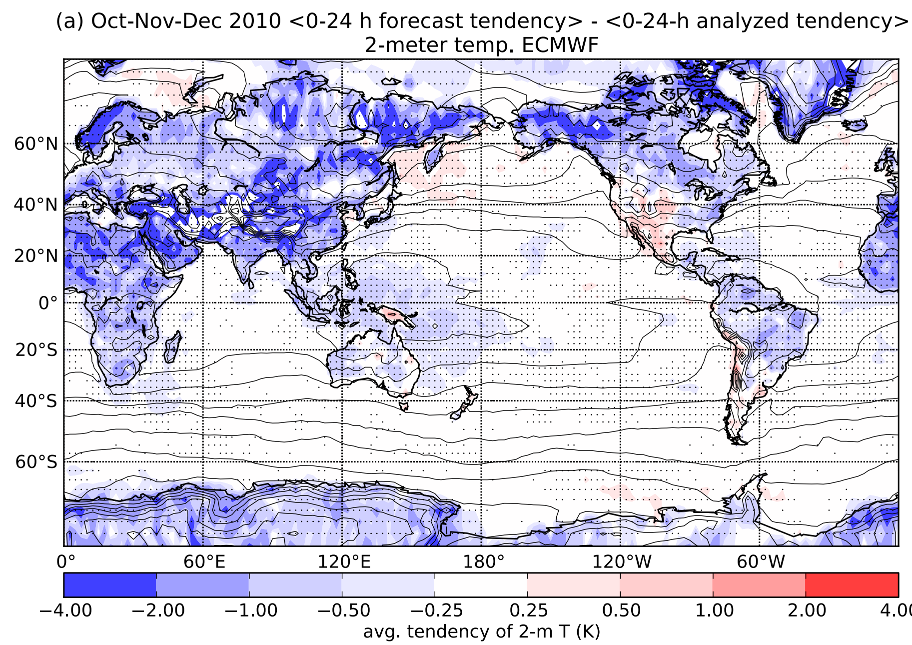
# Average 0-24 h tendencies in **2-m temp** <dF/dt> - <dA/dt>, Oct-Nov-Dec 2010

The plots show 3-month averaged forecast minus 3-month averaged observed tendencies, which hopefully are a reasonable diagnostic of day +1 model drift. Here, NCEP consistently exhibits positive tendencies over land, ECMWF negative tendencies over land. Let's focus on NCEP. Dots convey statistical significance.

NCEP




ECMWF



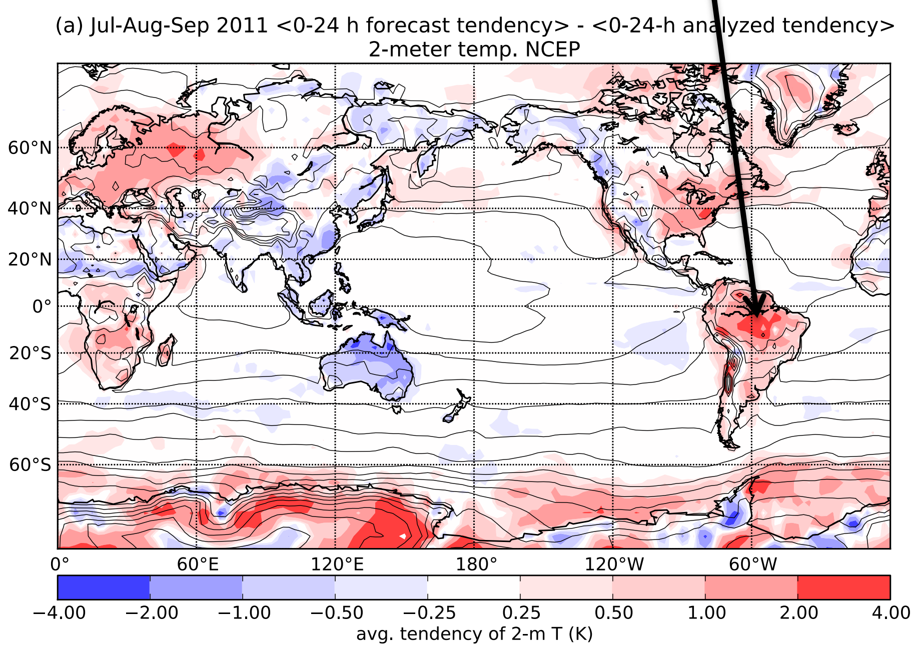


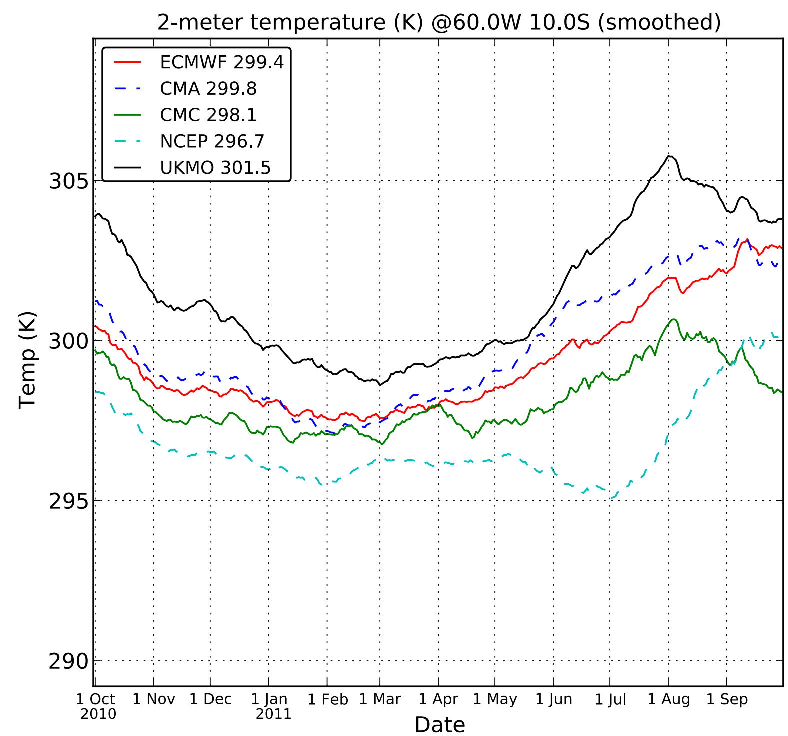

Note: other season's plots in supplementary slides, after conclusions.

# Average 0-24 h tendencies in NCEP **2-m temp** <F> - <A>, Jul-Aug-Sep 2011

What's going on here in the Amazon is rather puzzling. The 24-h forecasts are > 2K warmer on average than the analyses, which would lead one to believe that in this relatively data sparse region the analysis might thus have a warm bias with respect to the other analyses. Is this some case where there may be a repeated inappropriate negative (cooling) analysis increment in the data assimilation, and the 24-h forecast "snaps back" to a more reasonable climatology?

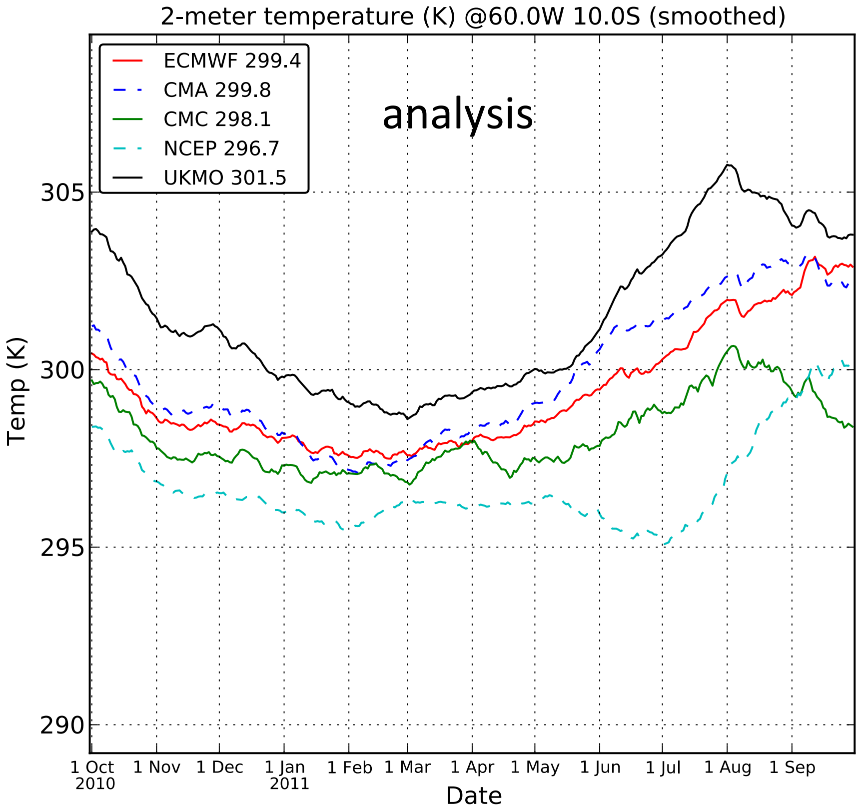
NCEP

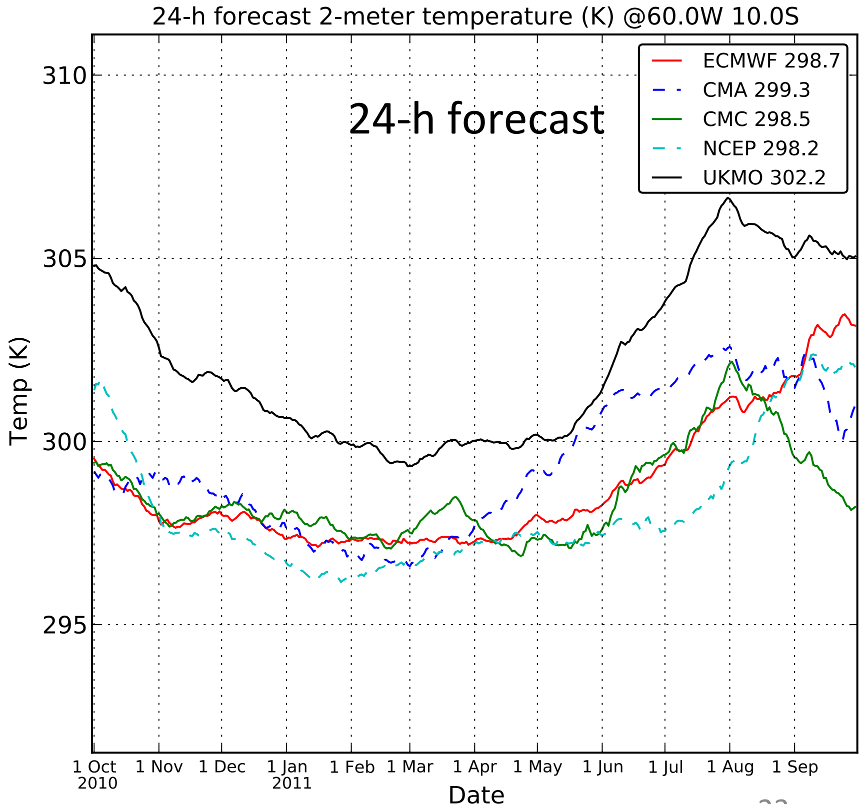


# Time series of analyzed and 24-h forecast 2-m temp (smoothed) in the Amazon

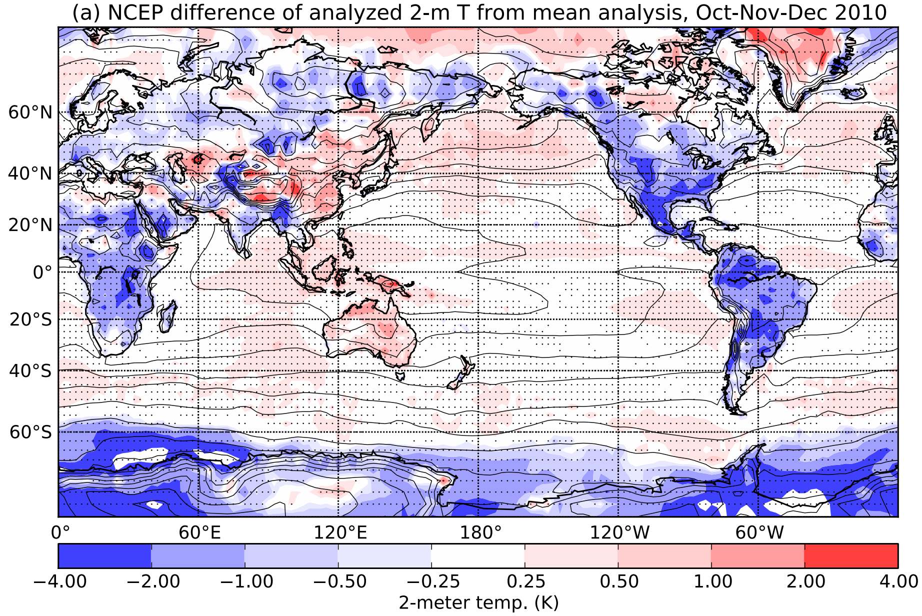
This suggests that perhaps for this location, since NCEP's 24-h forecasts become more consistent with the others while UKMO's become even more of an outlier, NCEP's problem is more likely due to some issues with the assimilation, while UKMO's problem is a forecast model bias problem.



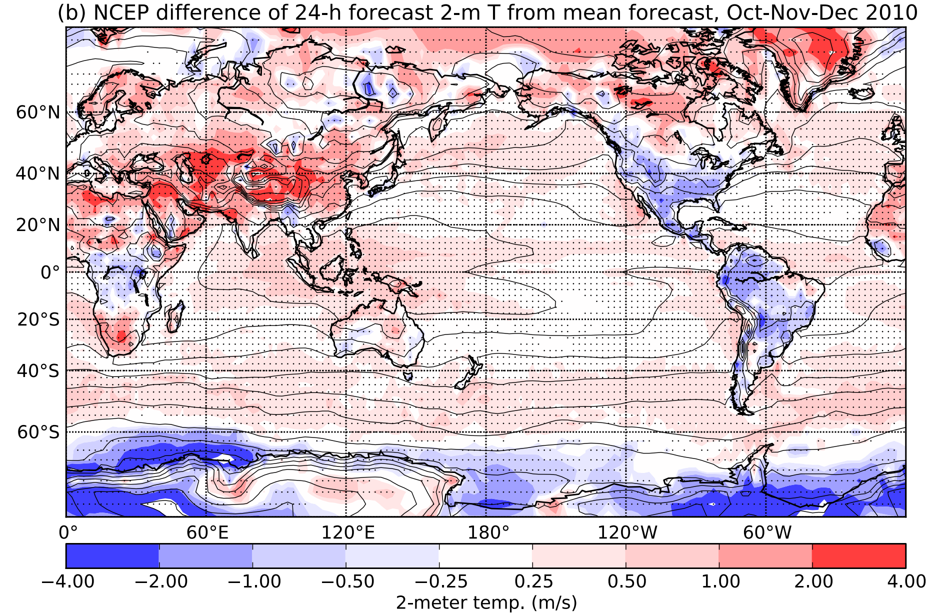



# Differences of NCEP **analyses and 24-h forecasts** from multi-center mean

analysis




24-h forecast

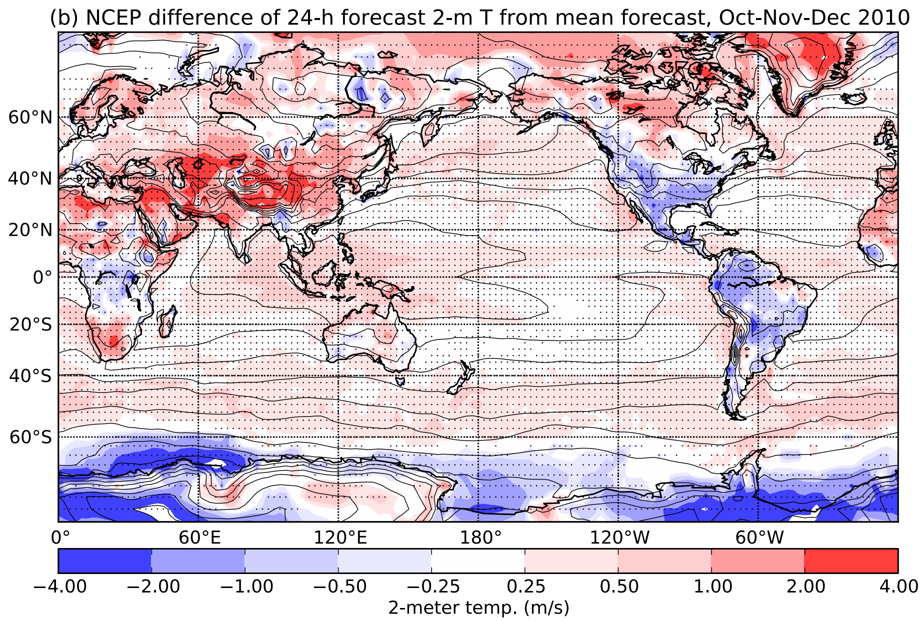



Generally NCEP land 2-m temperature analyses are cooler than other analyses, but by 24-h this is reversed and the NCEP forecasts are warmer in many locations.

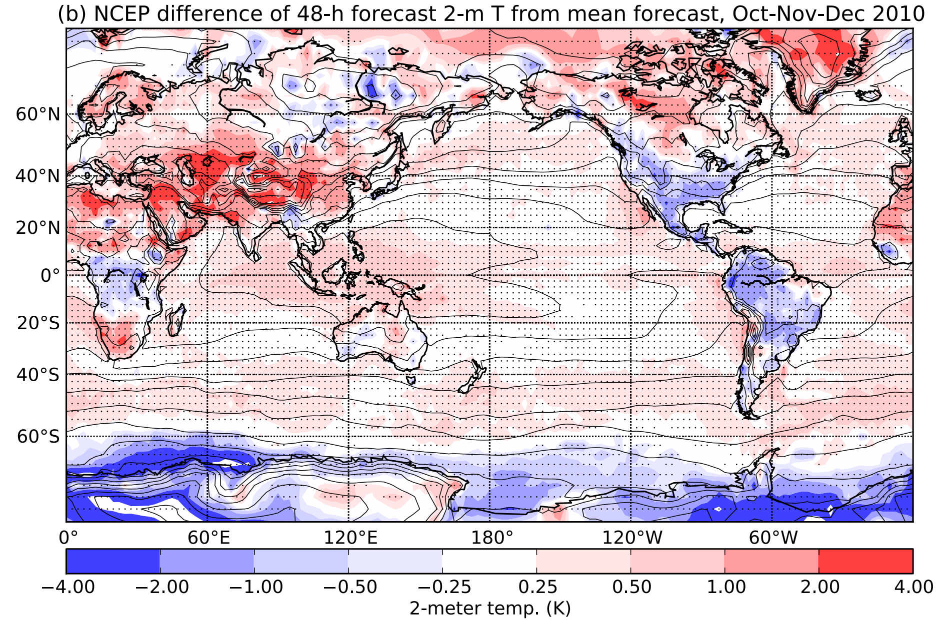
Is the model restoring itself from some disequilibrium during the analysis process, or are we seeing a general forecast bias?  Let's look at some longer leads.

# Differences of **24-h and 48-h forecasts** from multi-center mean

24 h




48 h

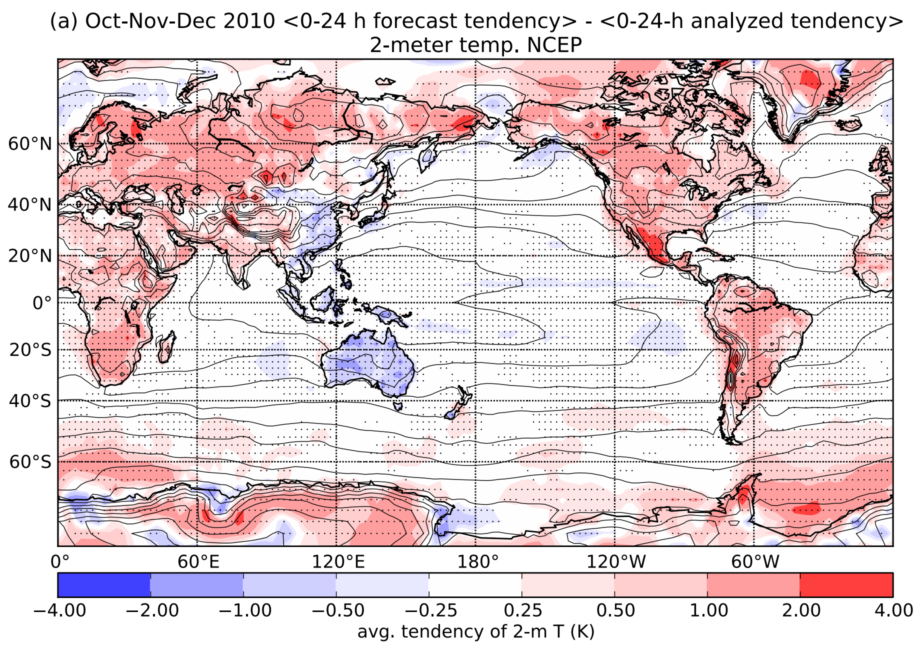



24-h and 48-h forecasts look **amazingly similar** for 2-m temp, suggesting that *the change in the first 24 h may be an adjustment due to inconsistencies introduced during the assimilation process.*

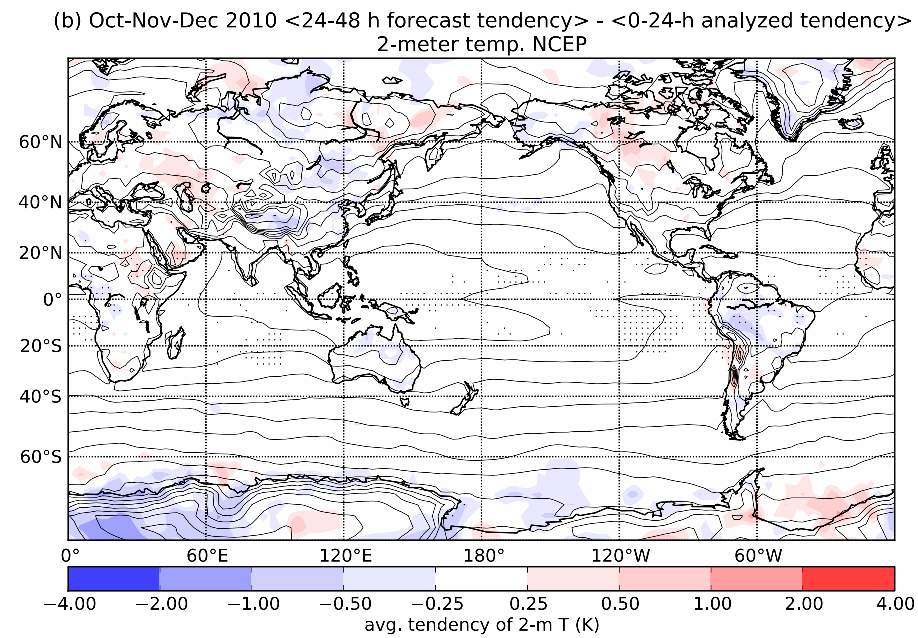
# NCEP 0-24 h and 24-48 h tendencies in **2-m temp** <dF/dt> - <dA/dt>

This confirms the 0-24 h tendencies for this field are transient adjustments.

0 to 24 h

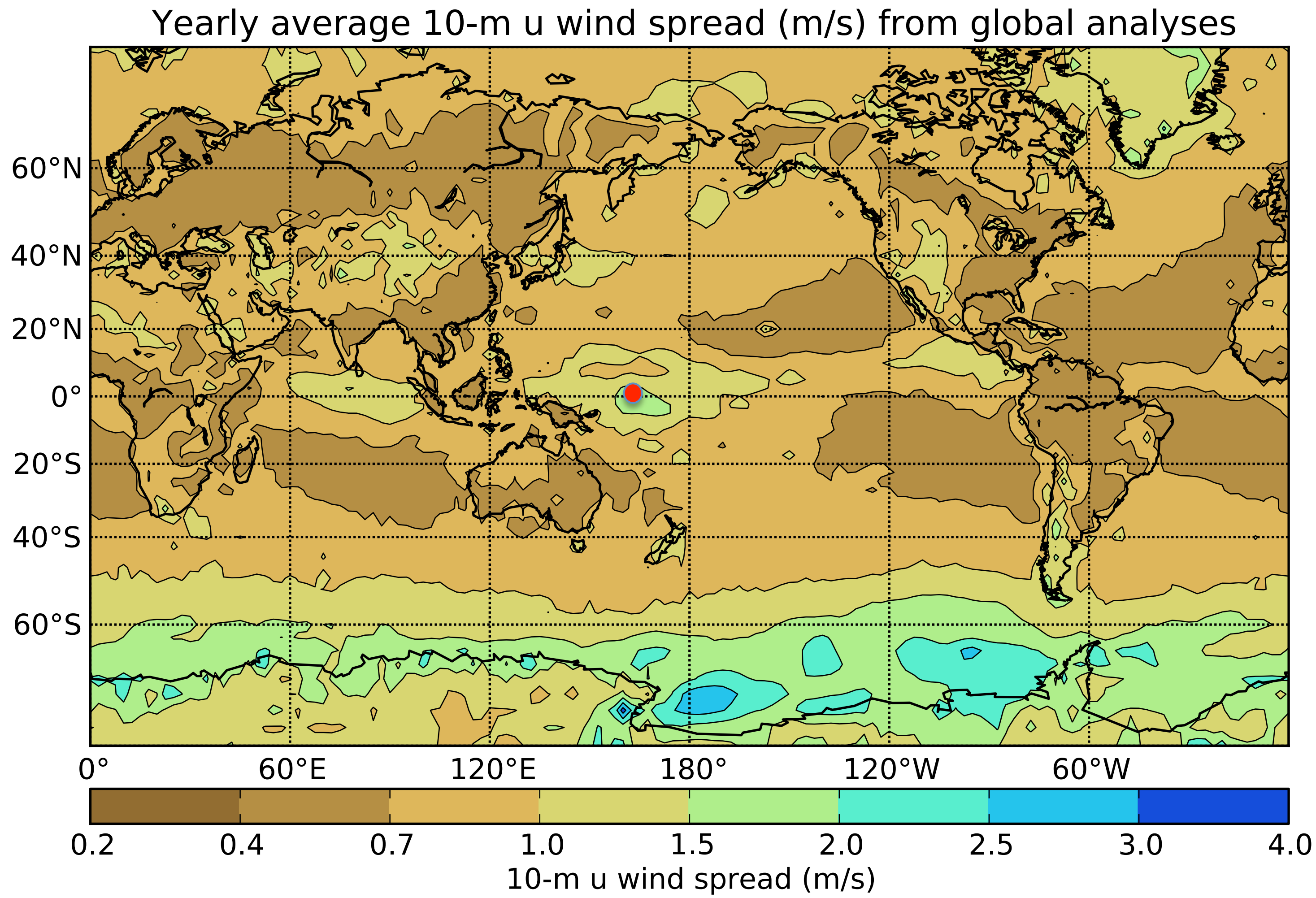



24 to 48 h

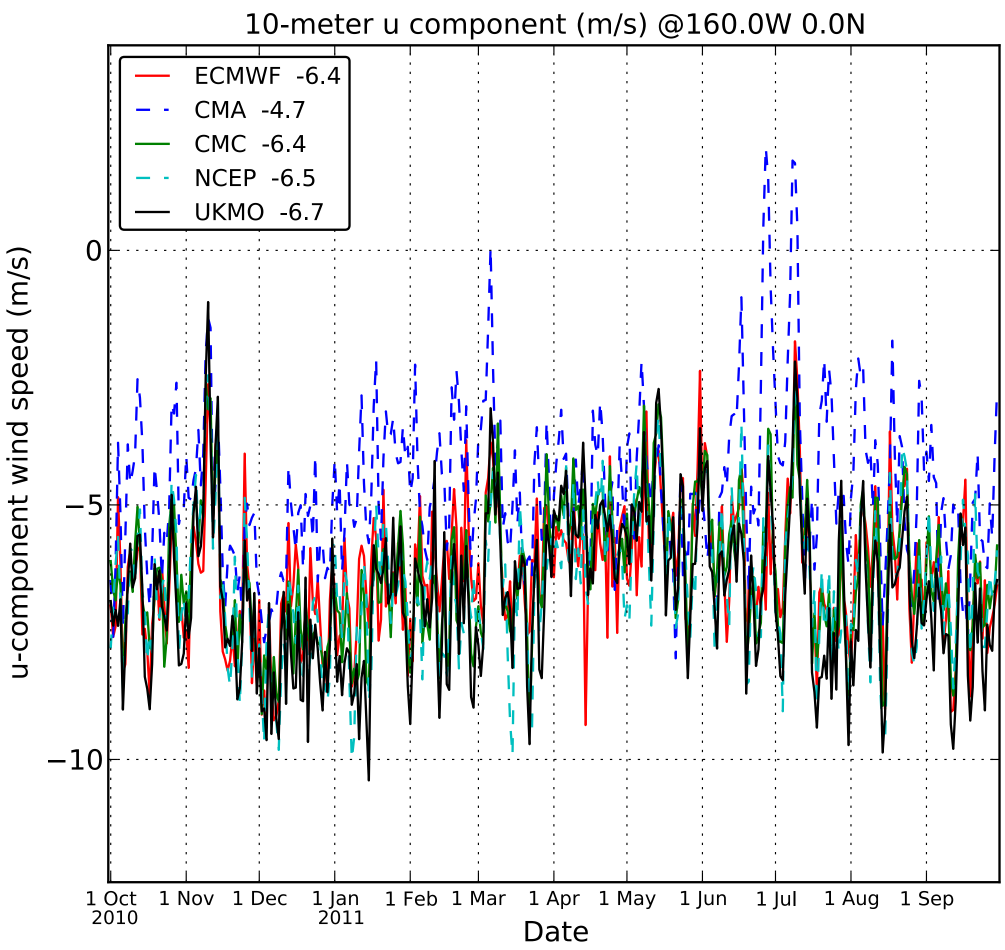


Let's look TIGGE analysis spread for other variables…

# Analysis spread, 10-m *u* component

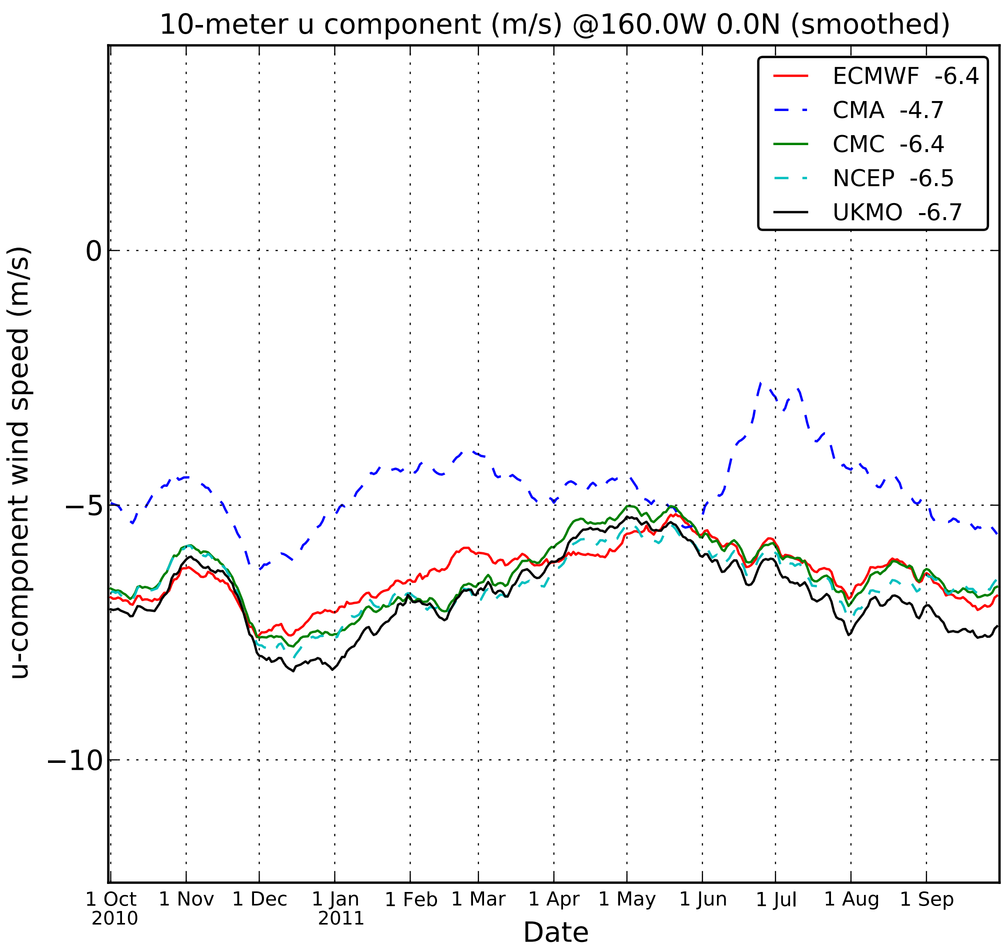


# Time series of 10-m u wind, equatorial west Pacific



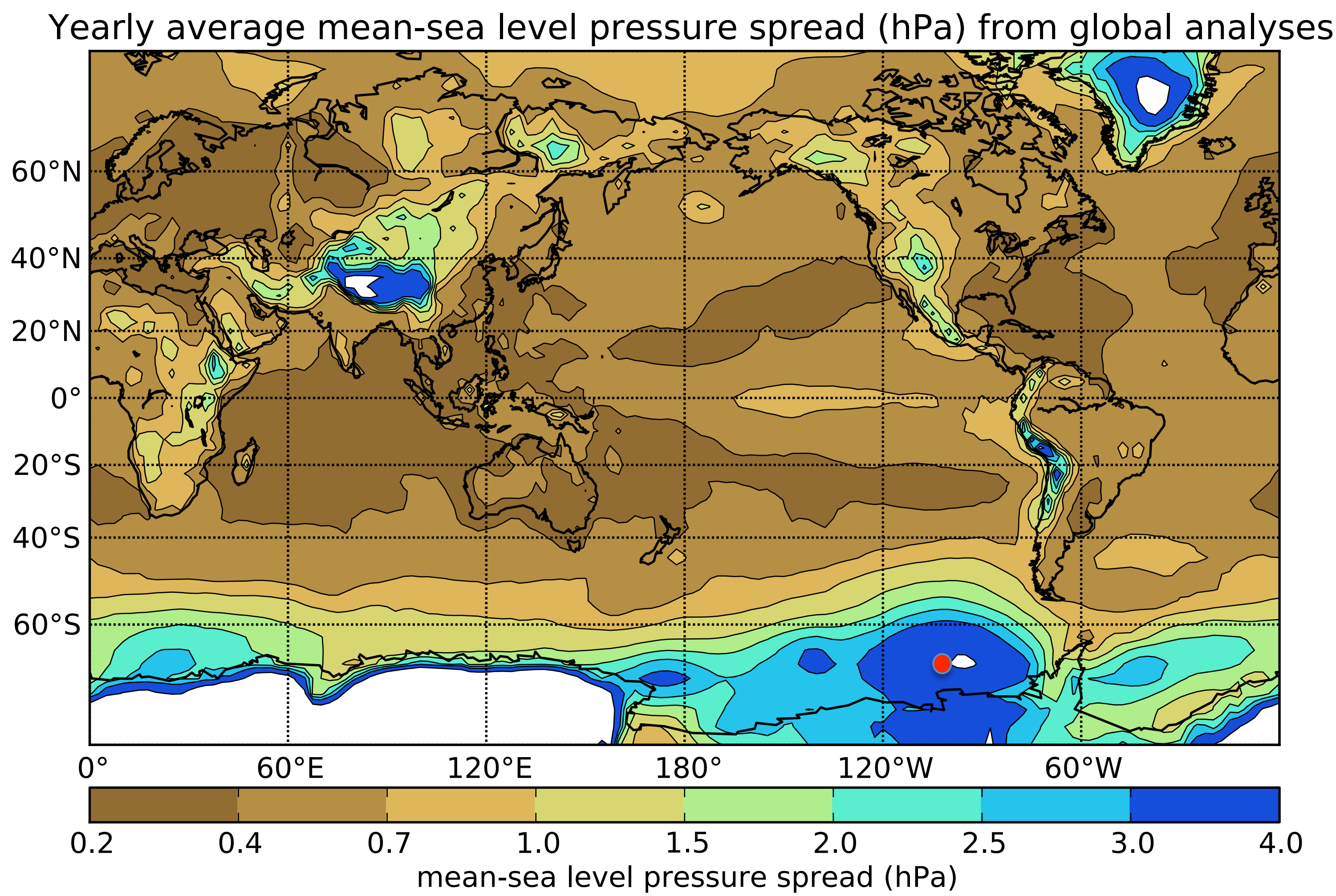

CMA analyses appear very different than the others in this region of critical ENSO and MJO variability.

# Time series of 10-m u wind, equatorial west Pacific, smoothed

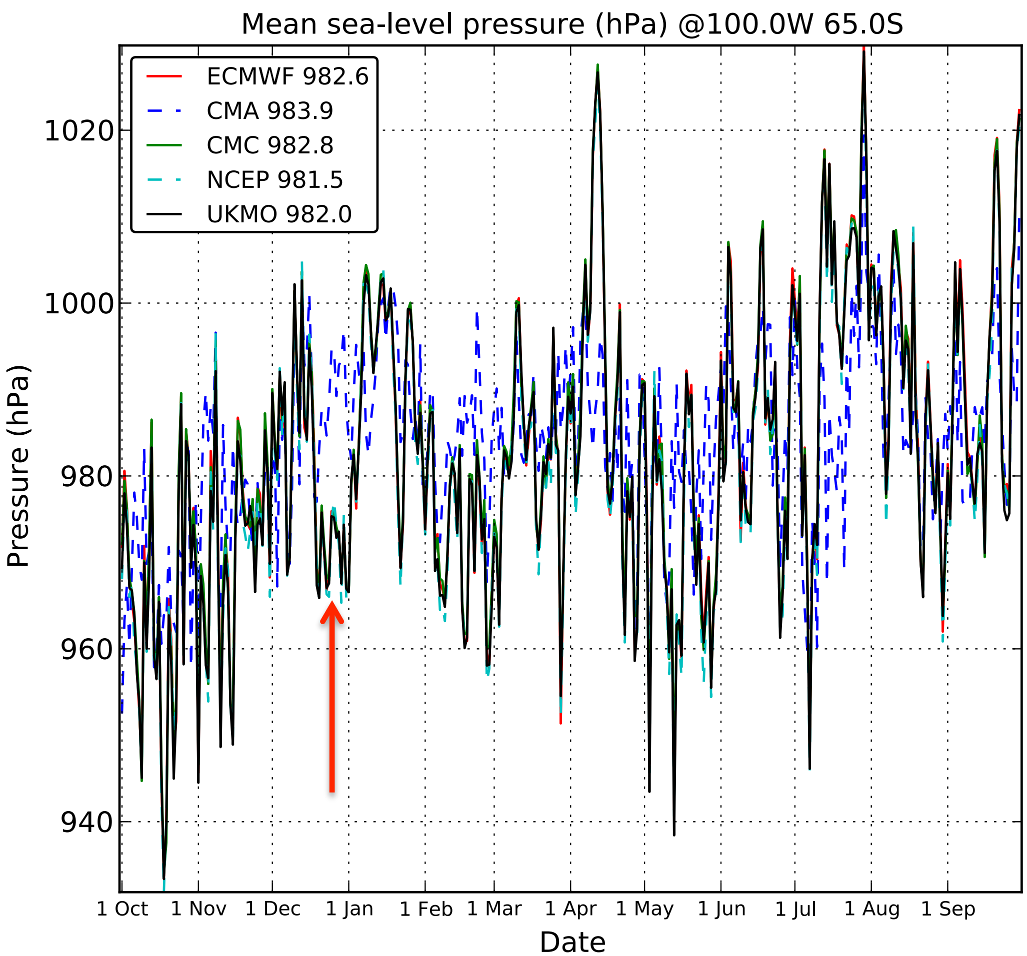



CMA analyses appear very different than the others in this region of critical ENSO and MJO variability.

# Analysis spread, mean sea-level pressure



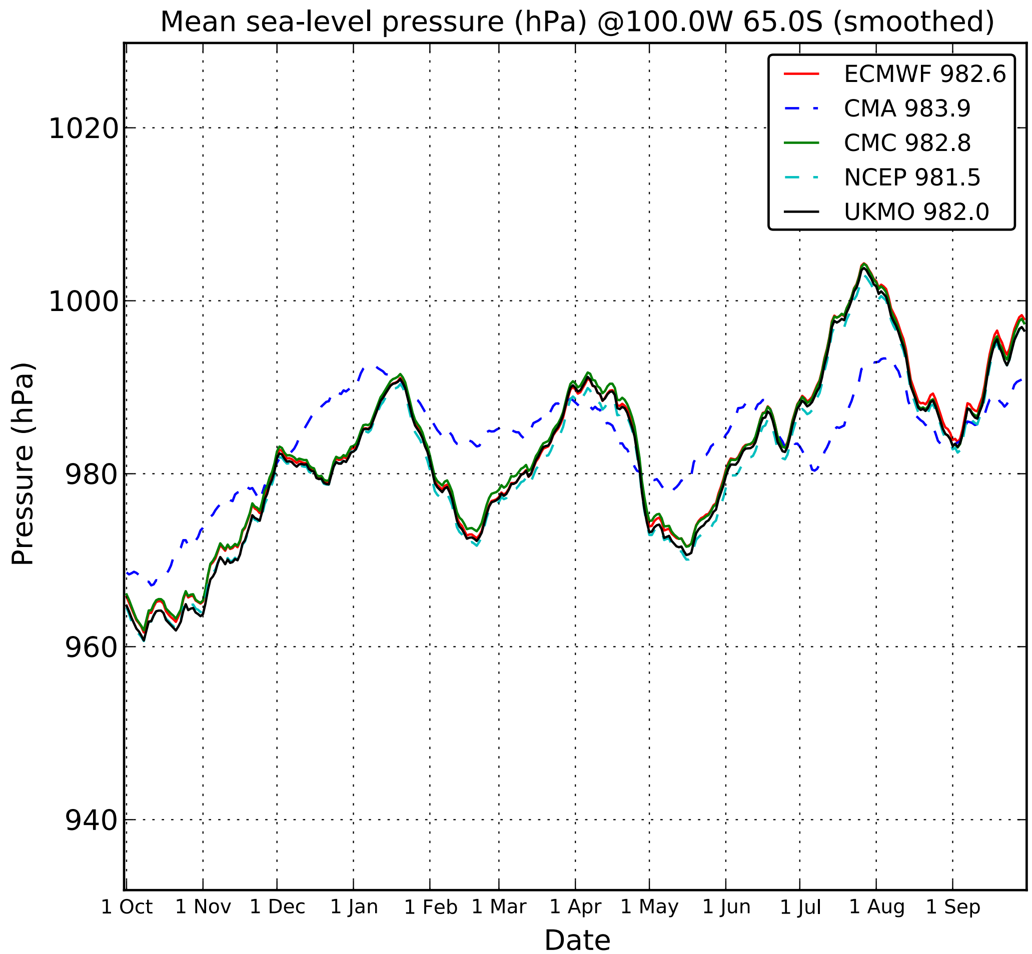
# Time series of SLP analyses, southern ocean




CMA analyses seem to have **a lot** less variability, not capturing the depth of the lows, strength of the highs here.

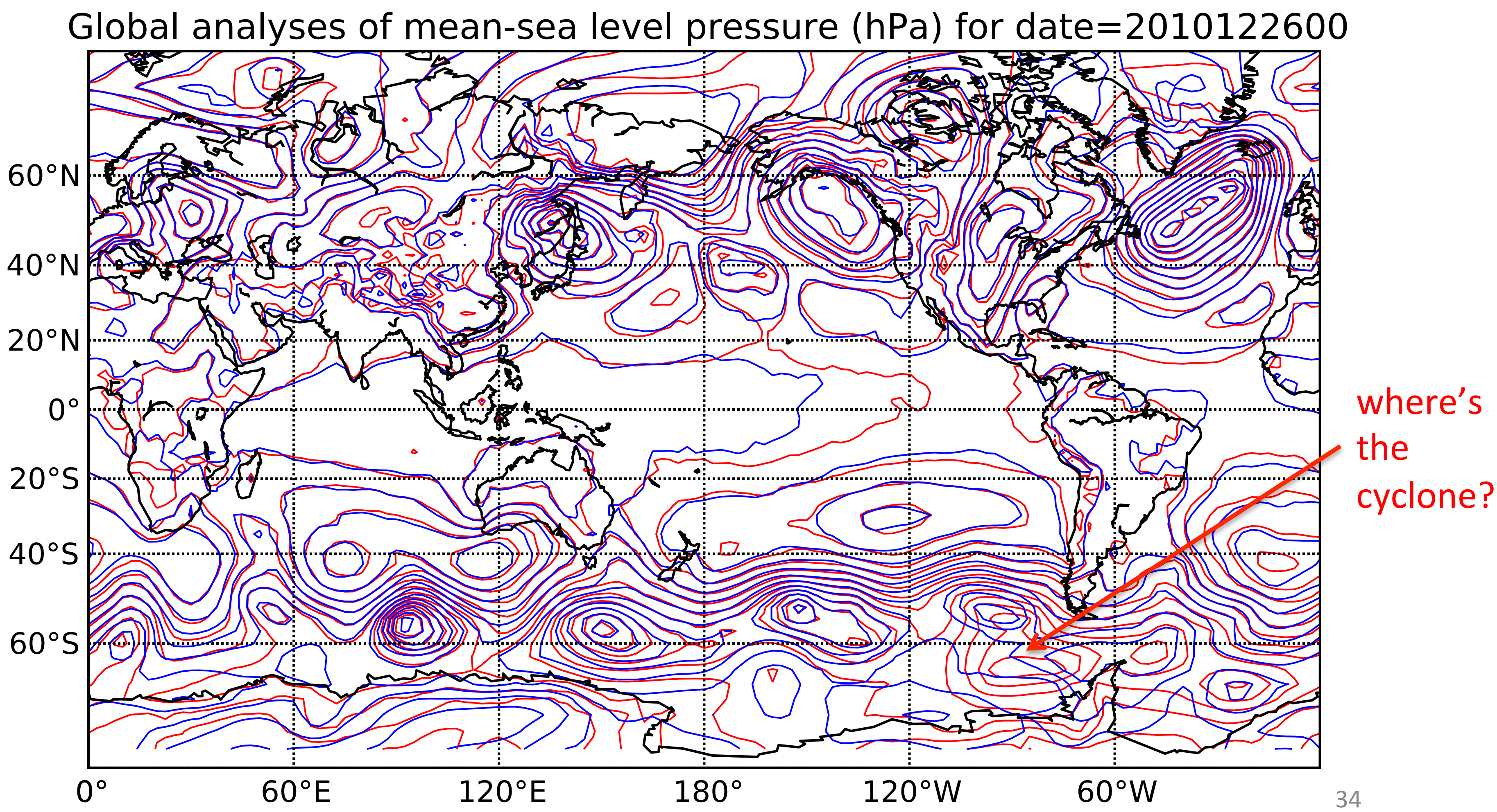
At dates indicated by red arrow, note that all other centers agree on a period of sustained low pressure of ~ 970 hPa, while CMA is ~ 990 hPa. 20 hPa different!

# Time series of SLP analyses, southern ocean (smoothed)

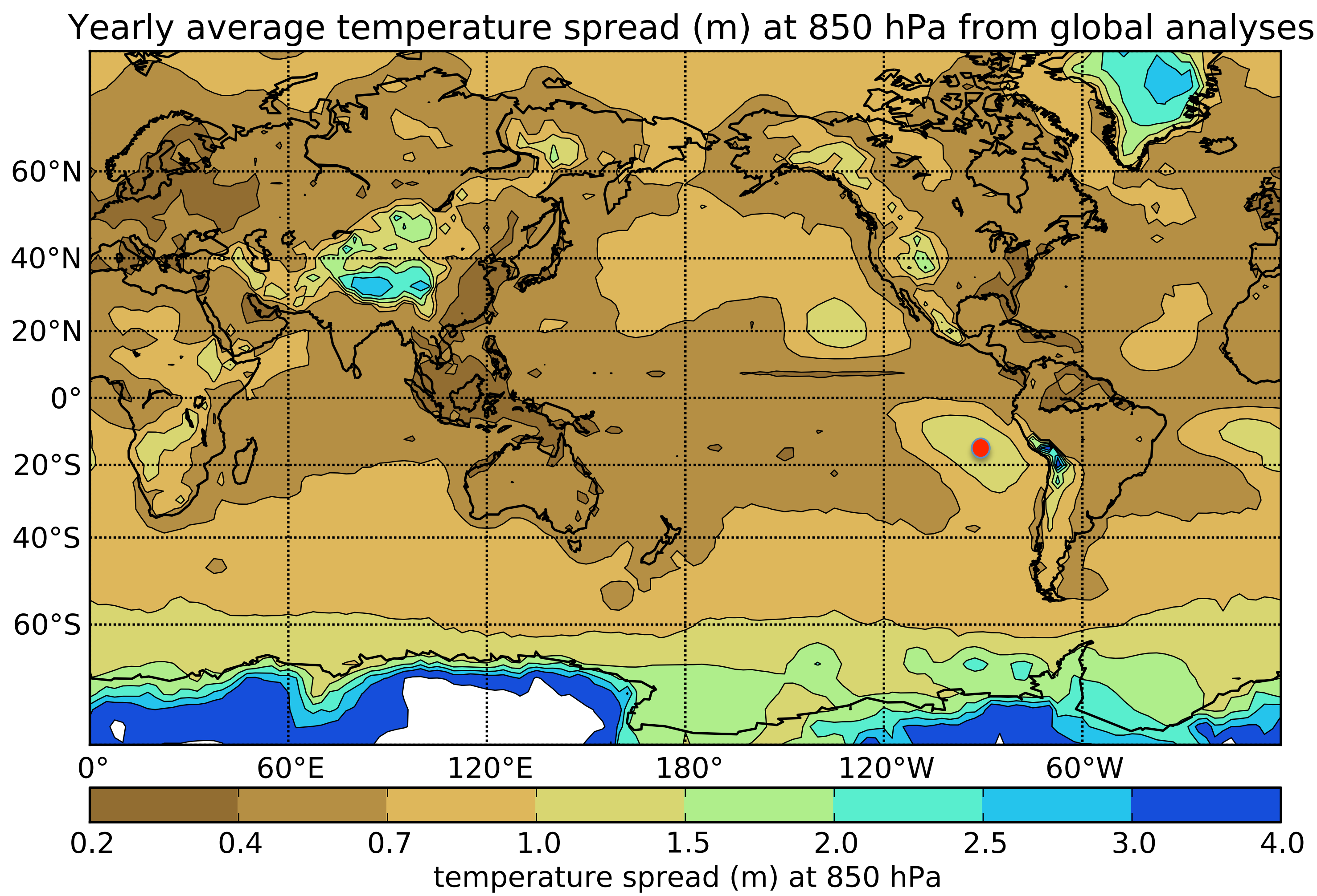



CMA's reduced variability is evident as well in the 30-day mean.

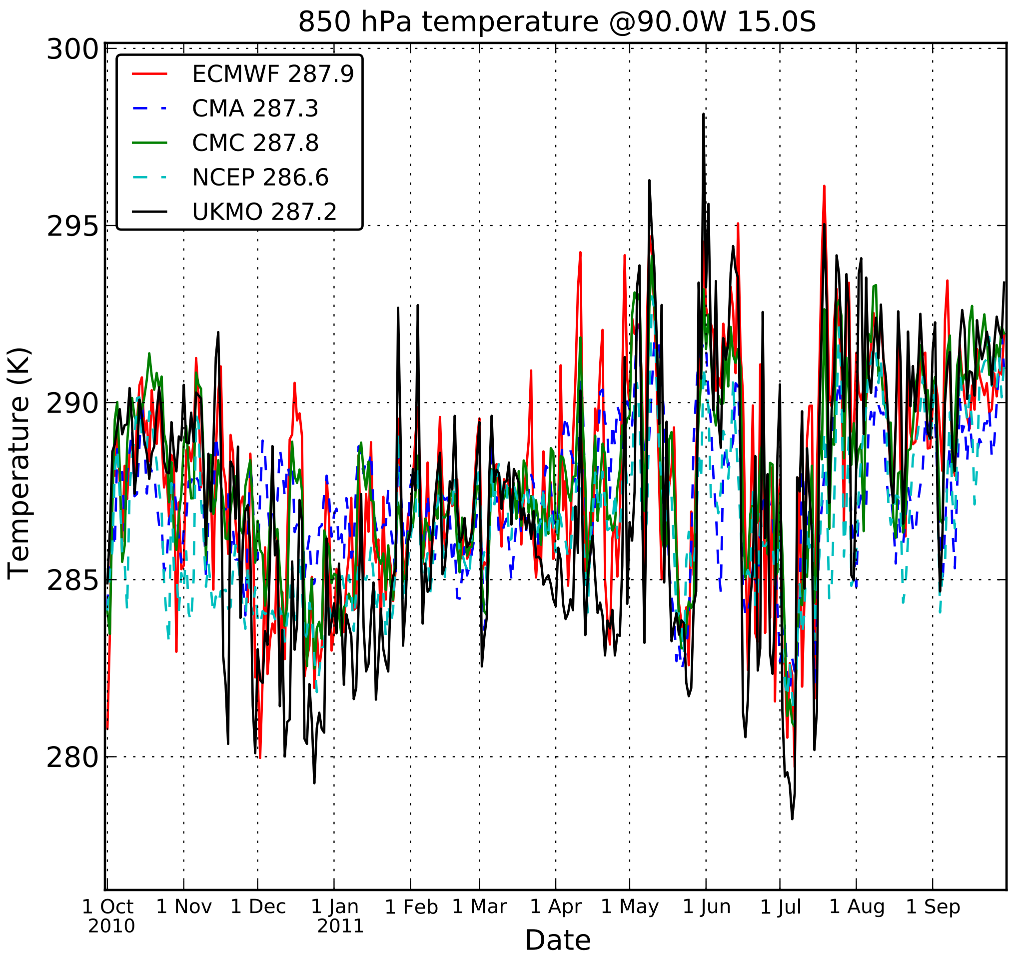
# Example: ECMWF (red) & CMA (blue) MSLP analyses, 2010/12/25/00Z



# Analysis spread, temperature at 850 hPa

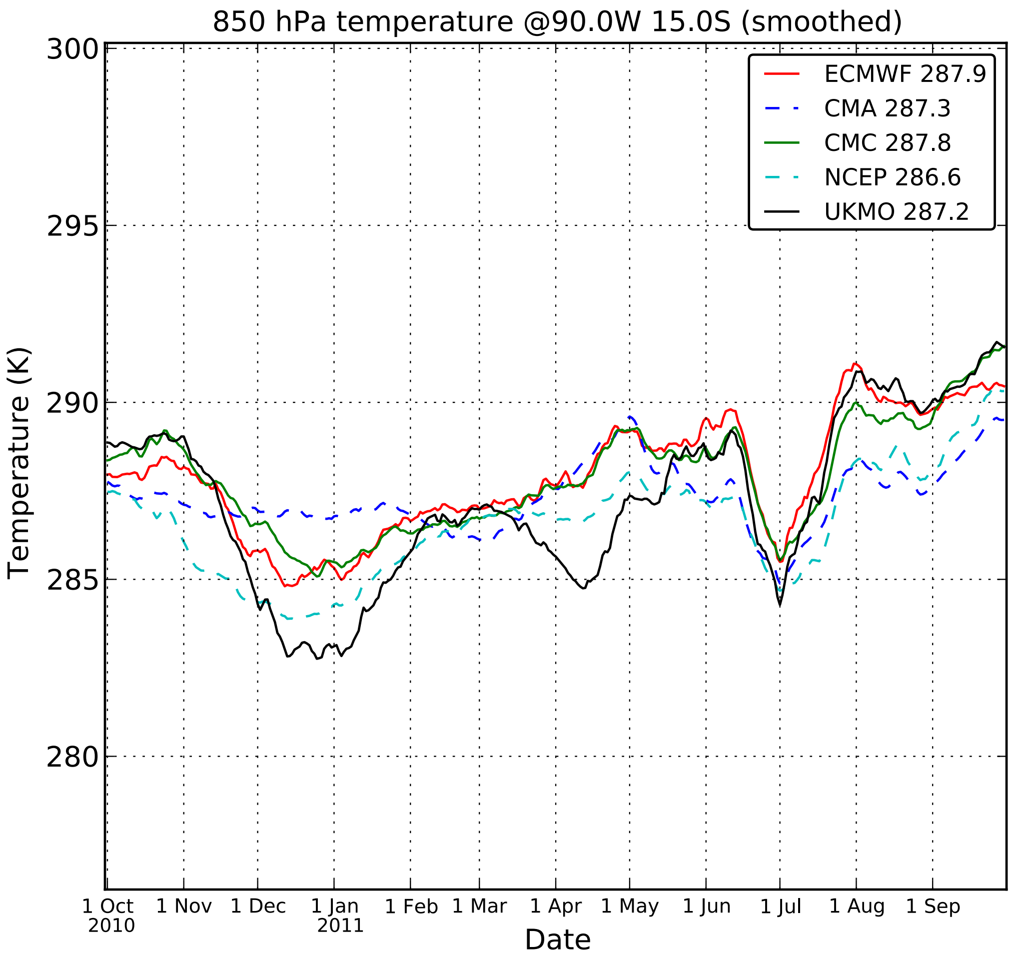


# Time series of 850 hPa temperature, south of Galapagos




Life in the subtropical ocean stratus deck. Differences appear more random. Probably there are some large differences in how the models parameterize stratus clouds and their radiative impacts.

# Time series of 850 hPa temperature, south of Galapagos (smoothed)



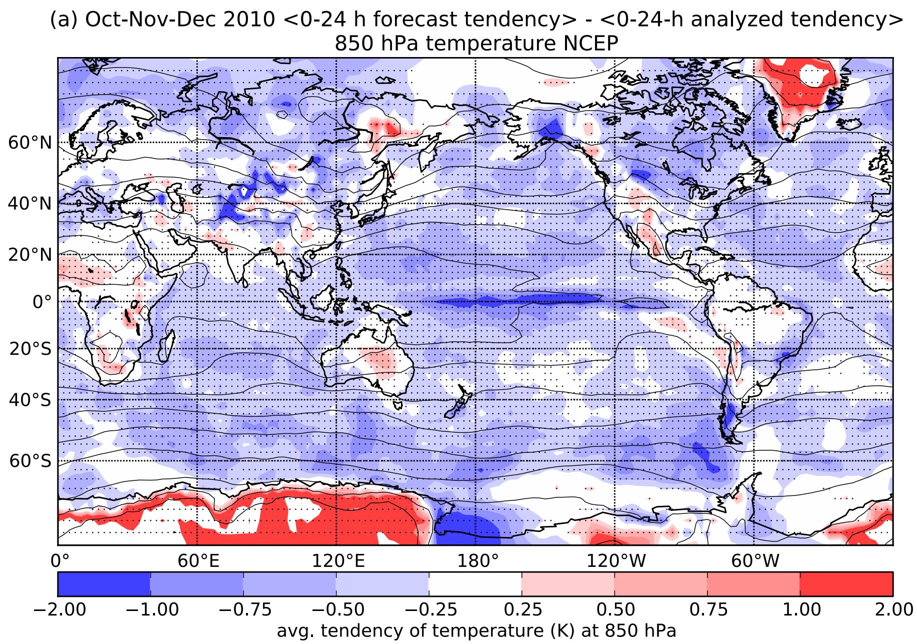

Again, CMA appears to have less temporal variability.

For several periods, UKMO is the outlier.

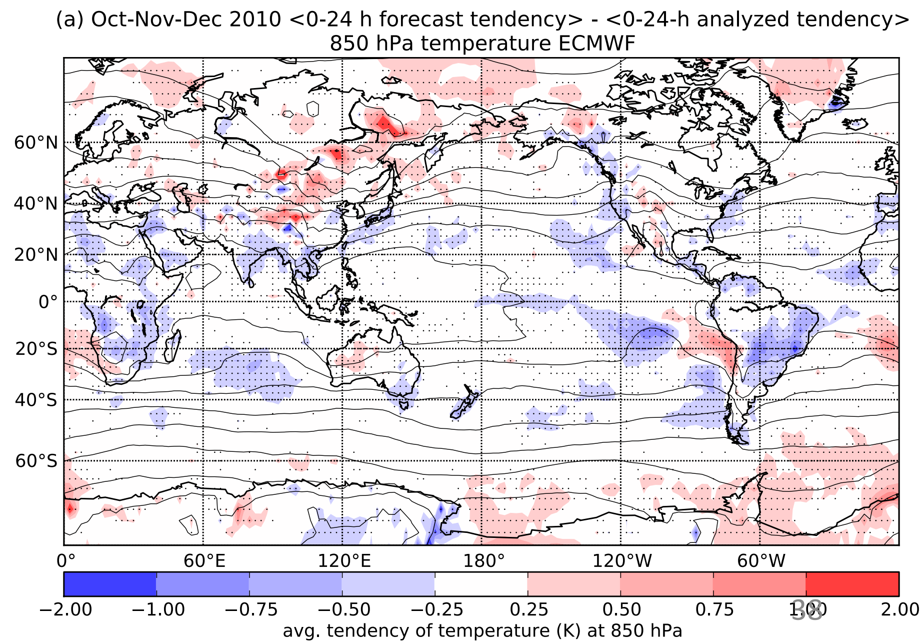
# Average 0-24 h tendencies in **850 T** <dF/dt> - <dA/dt>, Oct-Nov-Dec 2010

The systematic tendencies for NCEP are much larger than for ECMWF, and very large (>2K/day) over Antarctica, Greenland (but effectively below ground). Over the oceans, there are almost uniformly negative tendencies for NCEP in first 24h.

NCEP



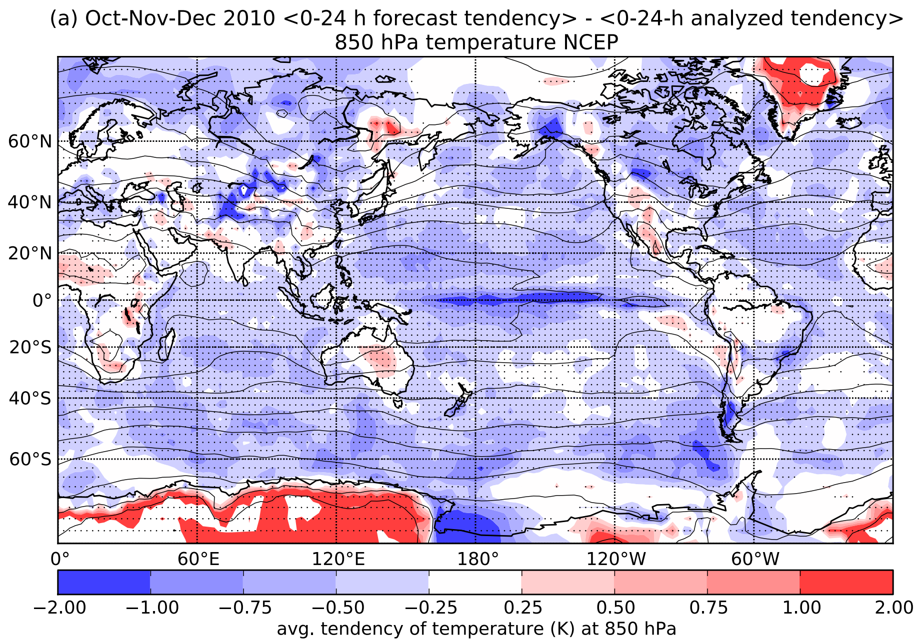

ECMWF



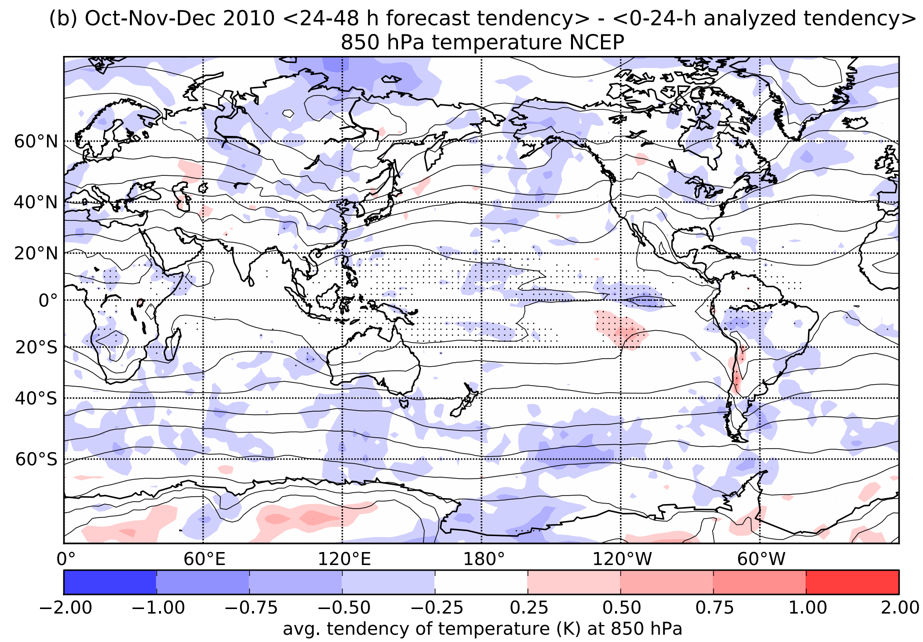
# NCEP 0-24 h and 24-48 h tendencies in **850 hPa temp** <dF/dt> - <dA/dt>

Again, NCEP 0-24 h tendencies for this field are dominated by transient adjustments, and the magnitude of the average of the tendencies is radically smaller on day 2 than day 1.

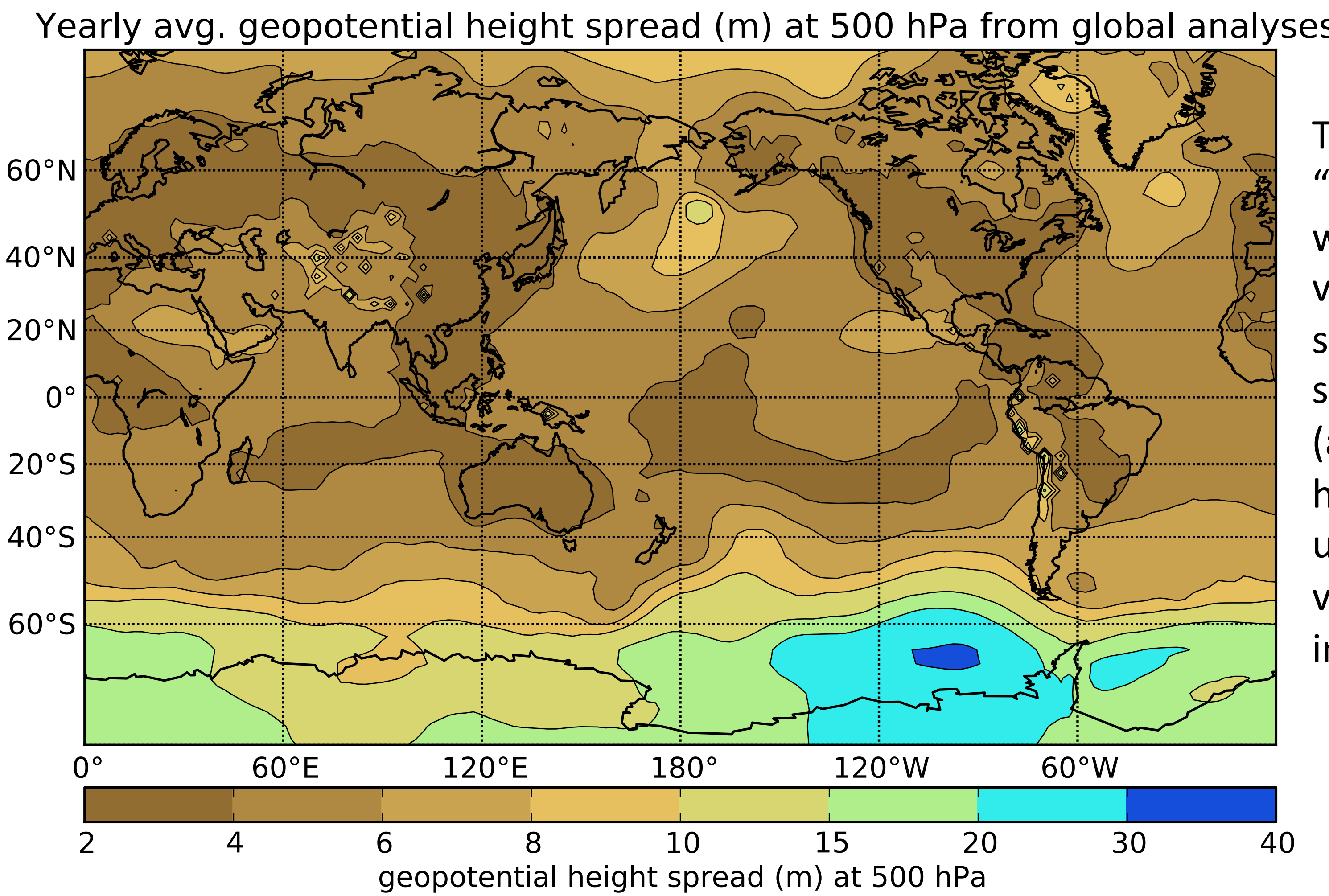
0 to 24 h




24 to 48 h



# Analysis spread, 500 hPa geopotential height

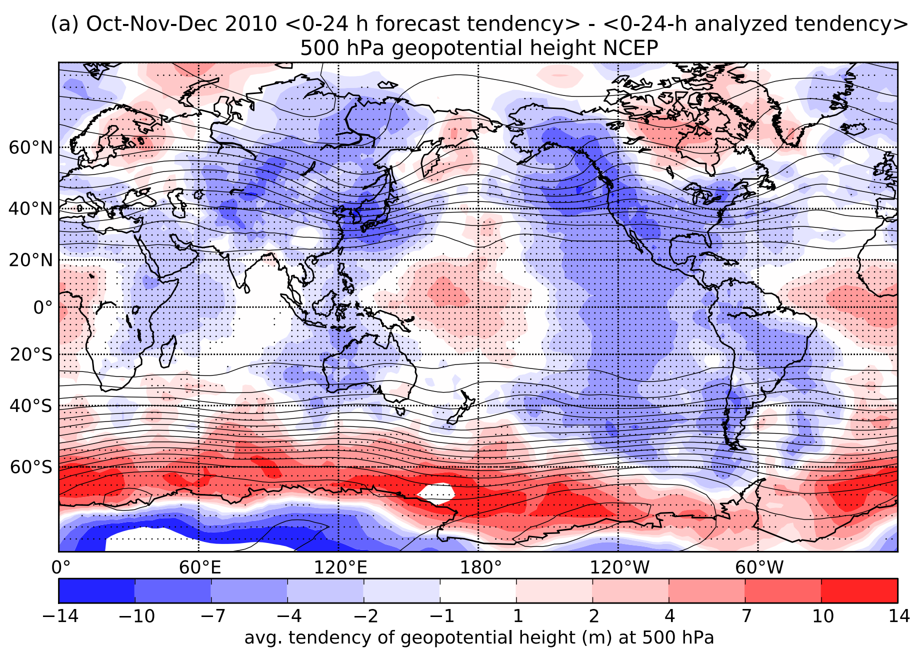



This looks more "well behaved", with more variability somewhat in the storm tracks (another sign of how 500Z is unlike many other variables of interest).

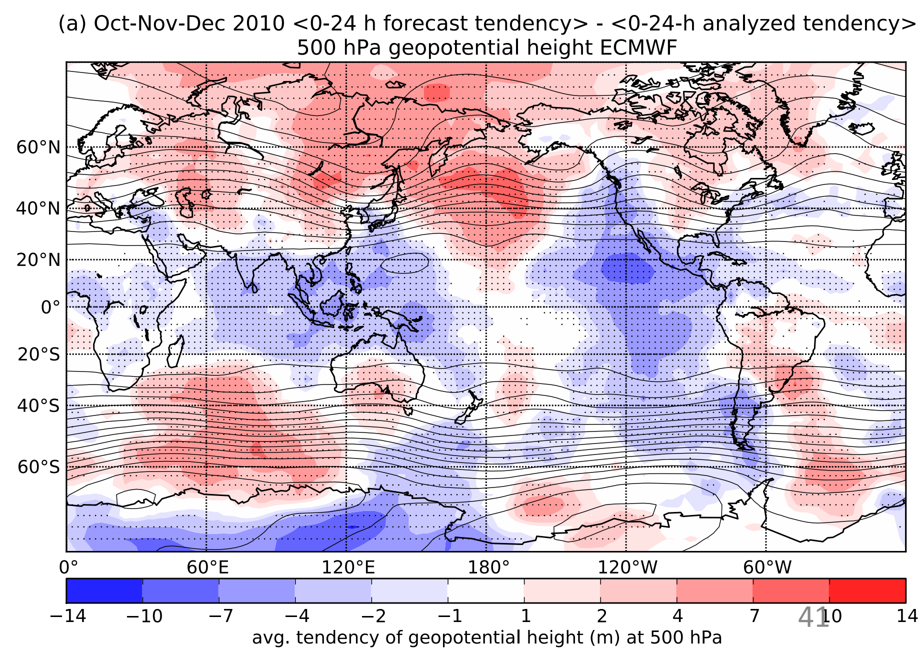
# Average 0-24 h tendencies in **500Z** <dF/dt> - <dA/dt>, Oct-Nov-Dec 2010

Again, very large tendencies over Antarctica for NCEP. Geopotential height falls there indicate, presumably, a strong anomalous cooling of air from surface to 500 hPa. Also: large height rises on poleward side of S. Hem. jet, indicating a systematic weakening of the jet.

NCEP



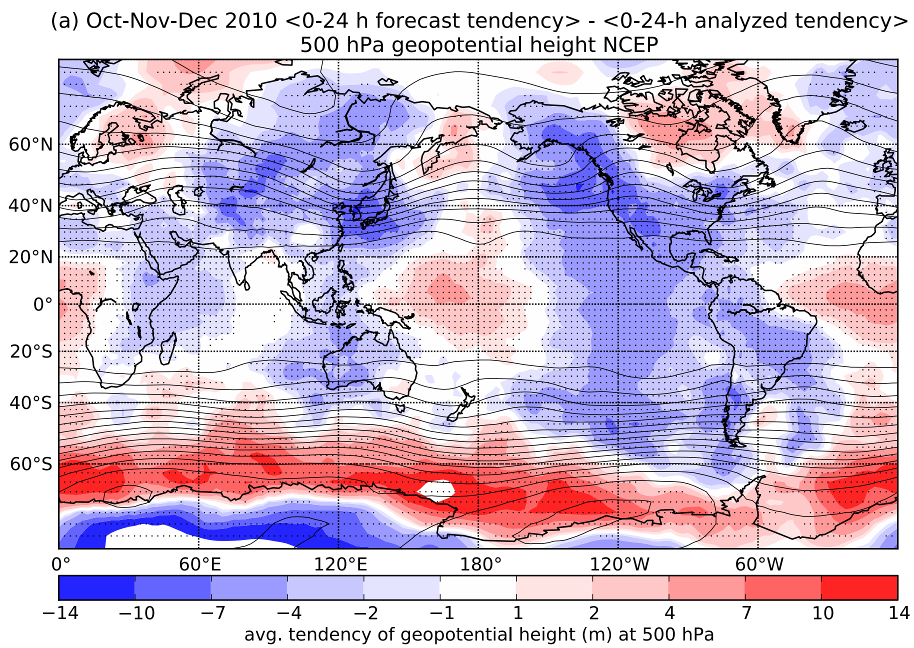

ECMWF



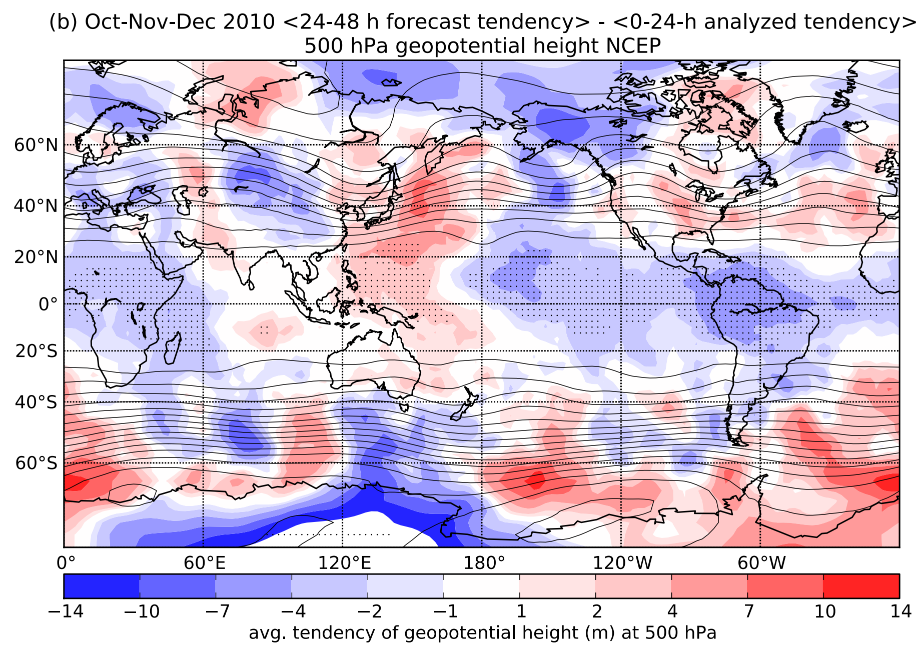
# Average 0-24 h and 24-48 h NCEP tendencies in **500Z** <dF/dt> - <dA/dt>

Consistent negative tendencies over Antarctica. Otherwise, the patterns show little similarity, and amplitudes of 24-48 h tendencies are comparable to those for 0-24 h. Also, note the lack of dots corresponding to many regions with intense colors. Despite relatively large changes, these are not statistically significant.

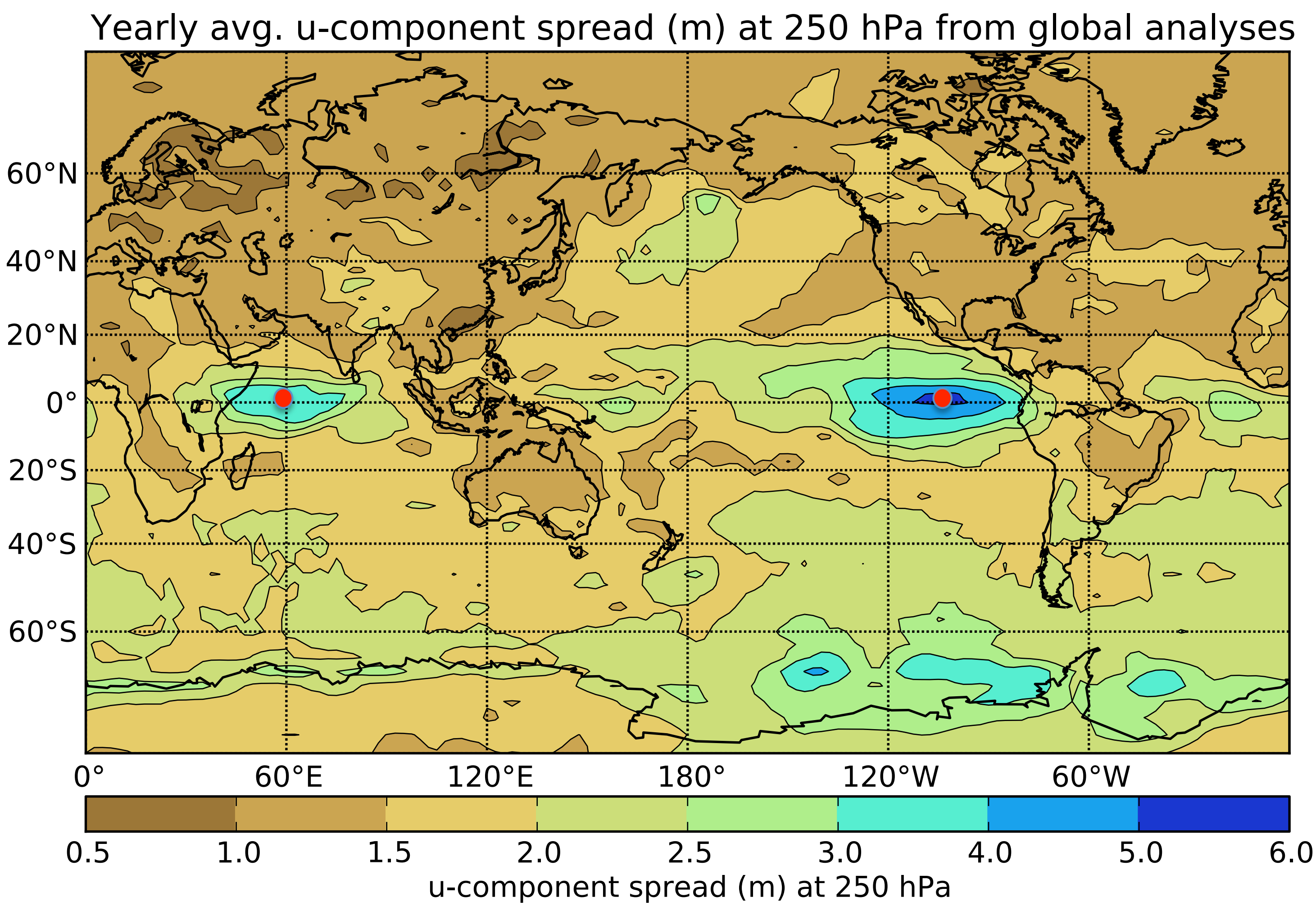
0-24 h




24-48 h

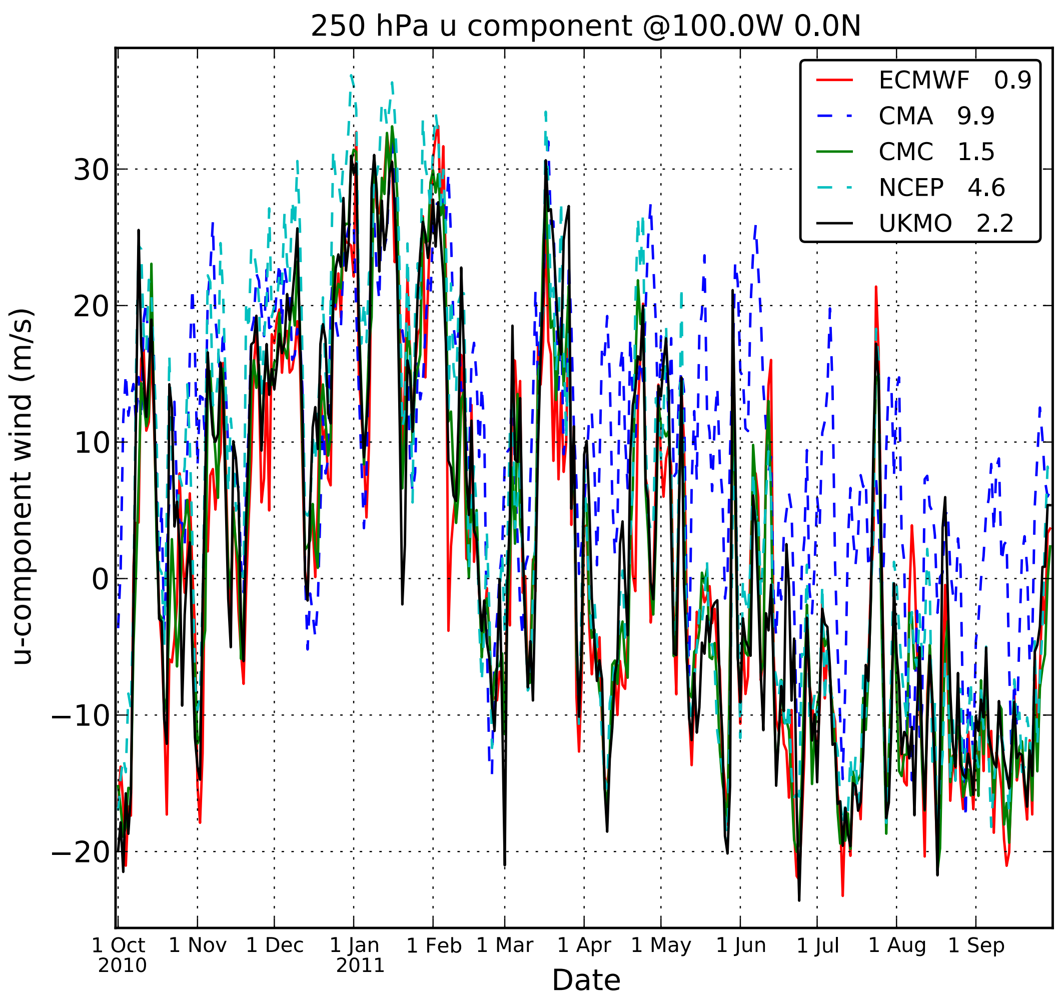


# Analysis spread, 250 hPa u component



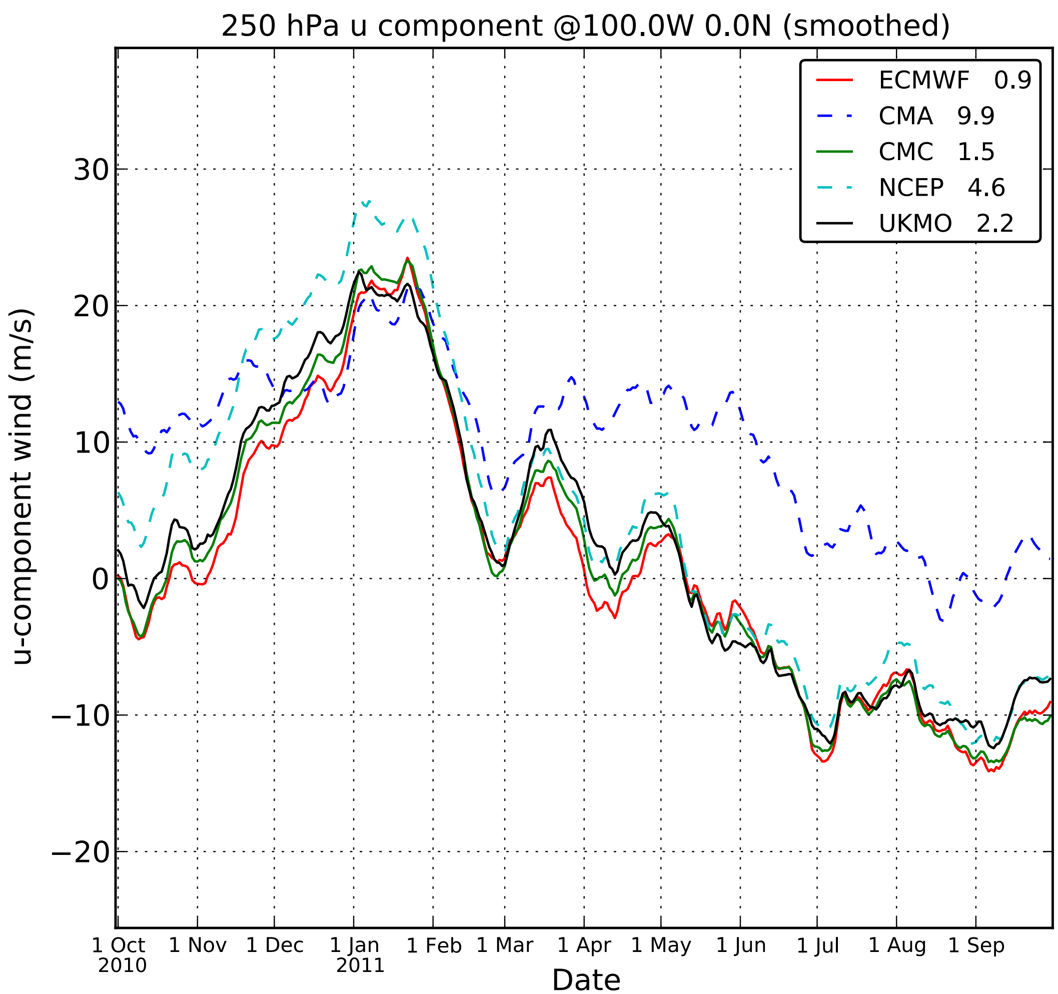

What's going on in tropical eastern Pacific and tropical western Indian Oceans?

# Time series of 250 hPa u-wind, eastern equatorial Pacific

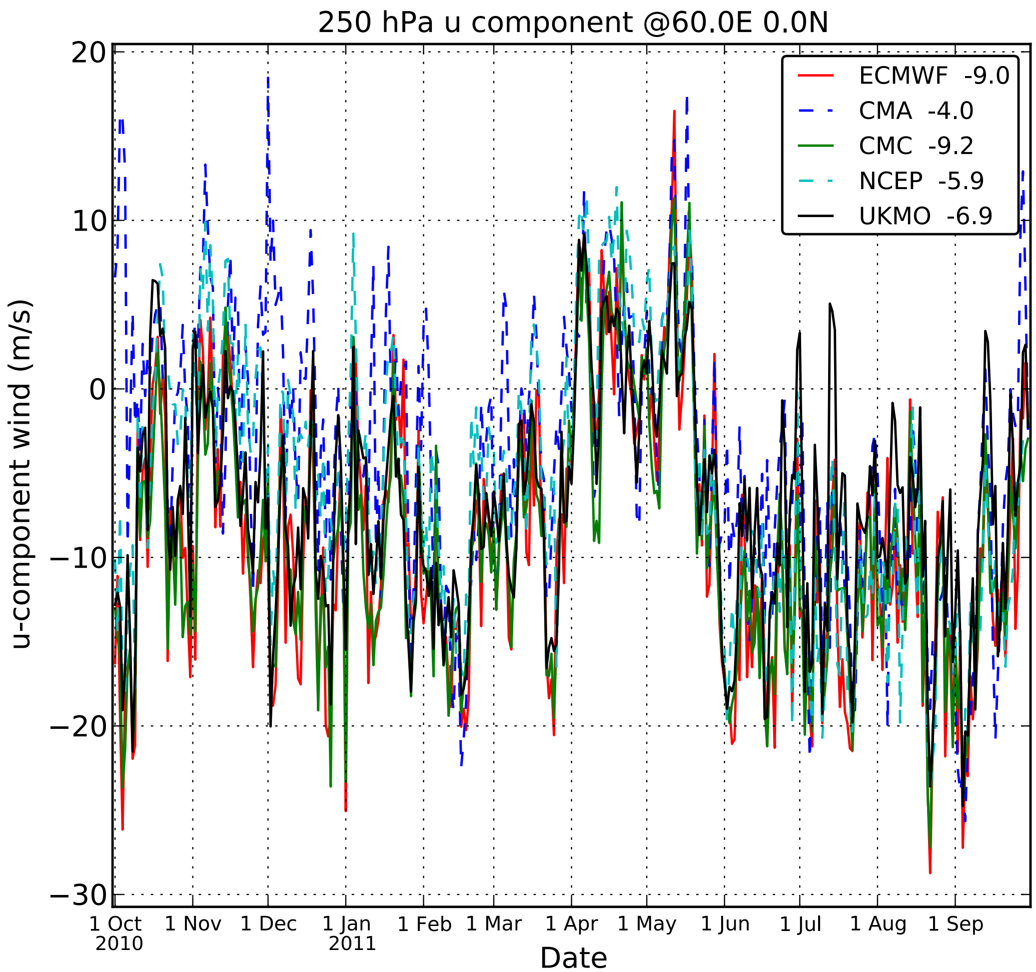



CMA, and to lesser extent NCEP, have much stronger westerly winds.

# Time series of 250 hPa u-wind, eastern equatorial Pacific (smoothed)

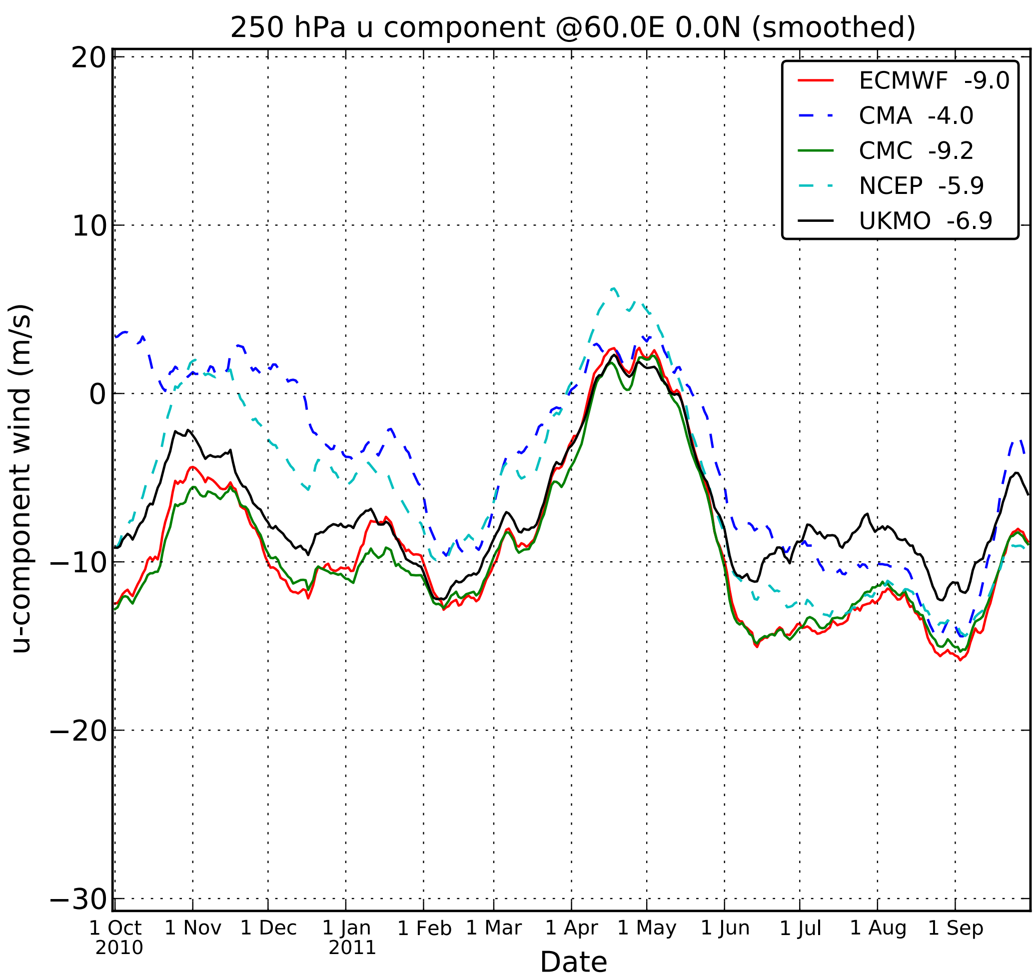


# Time series of 250 hPa u-wind, western equatorial Indian Ocean




CMA has much stronger westerly component, especially early during this period.

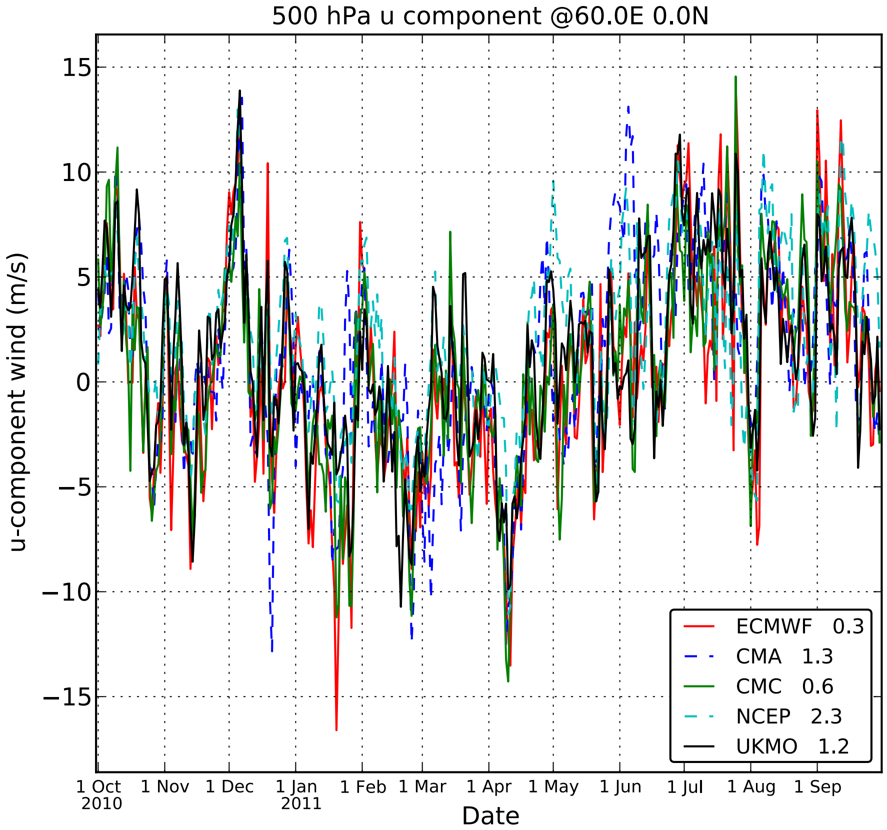
# Time series of 250 hPa u-wind, western equatorial Indian Ocean

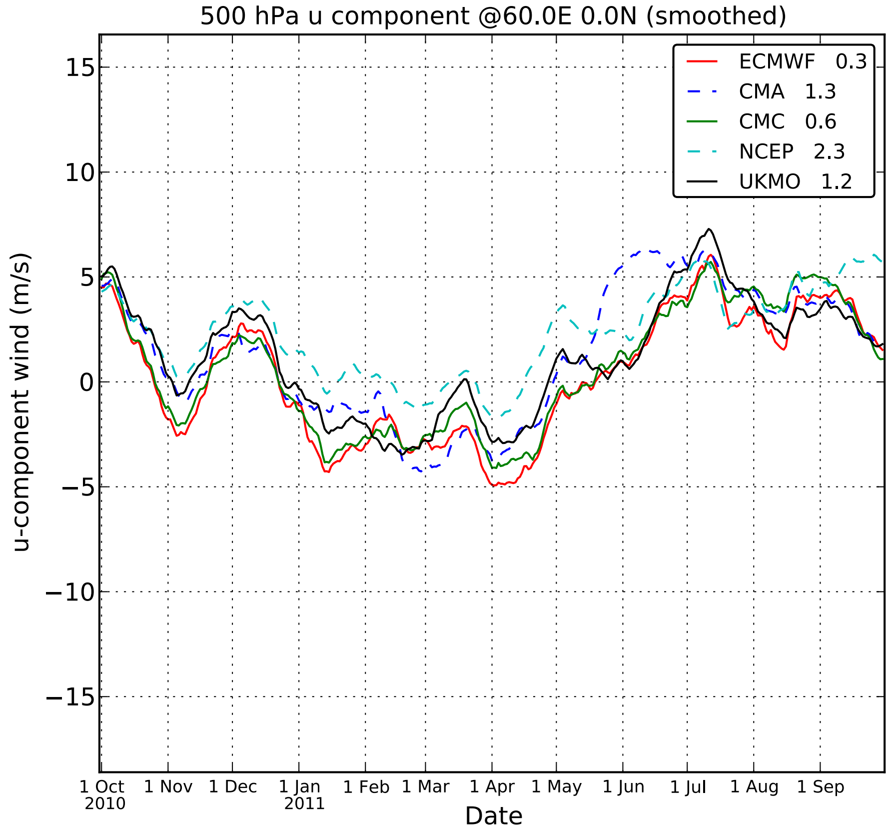



CMA has much stronger westerly component, especially early during this period.

It seems like these biases might have an effect on MJO.

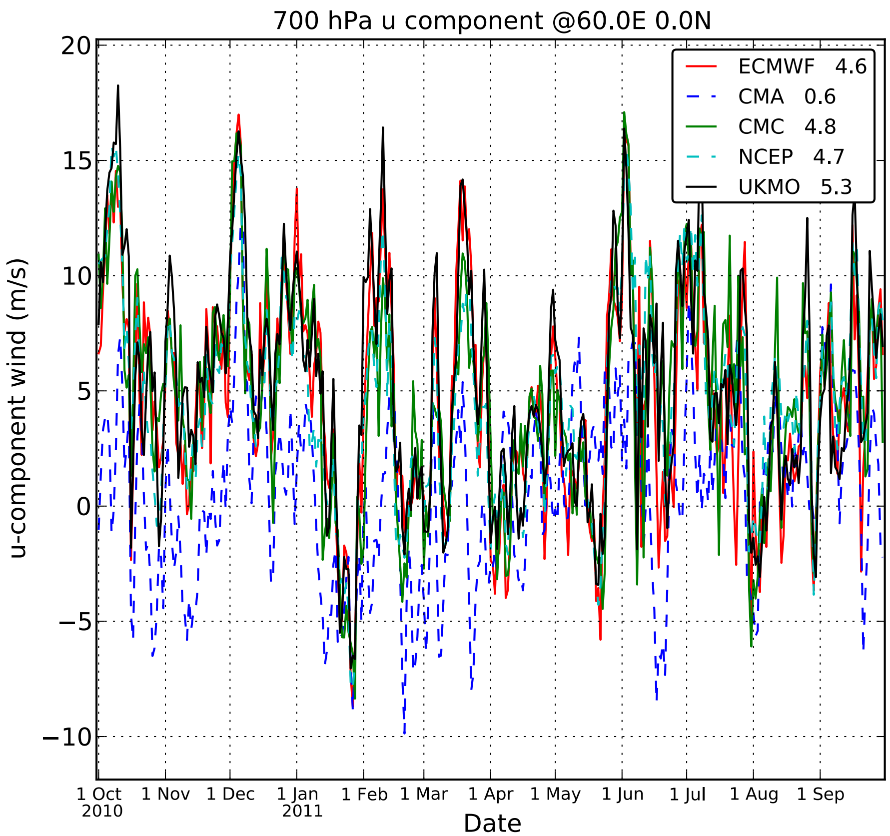
# Time series of 500 hPa u-wind, western equatorial Indian Ocean

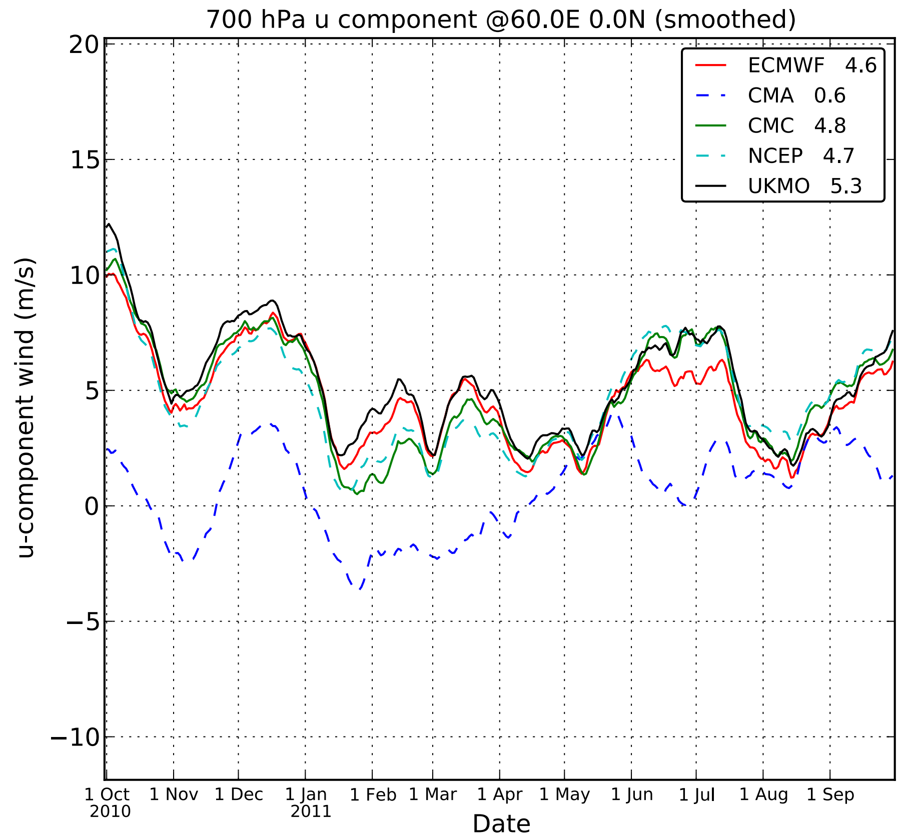



CMA is no longer the outlier here

# Time series of 700 hPa u-wind, western equatorial Indian Ocean




At 700 hPa, CMA is now consistently too easterly rather than westerly.
CMA may have some anomalous analyzed zonal circulation?

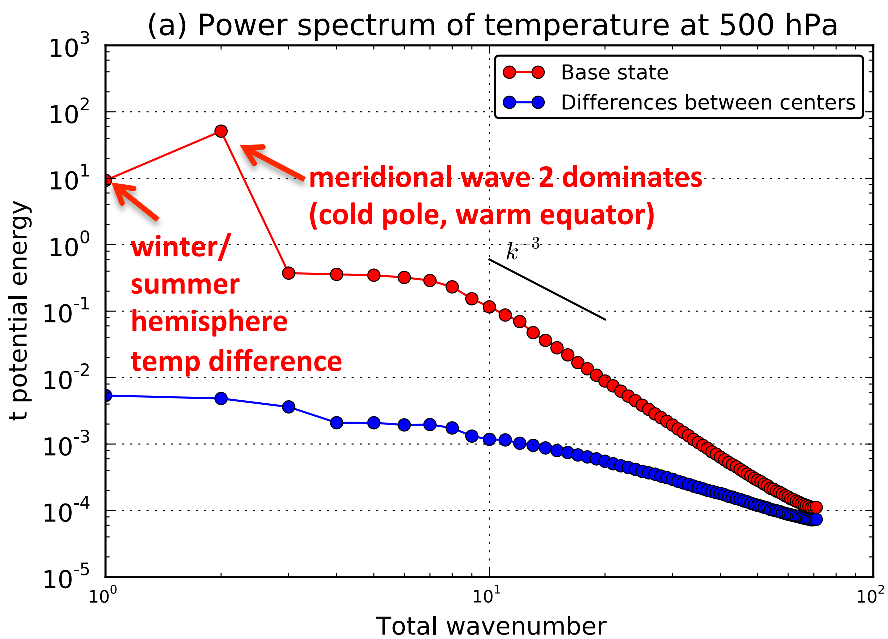
# Using TIGGE multi-center analyses to evaluate methods for initializing ensembles

- Theoretically, ensembles ought to sample the distribution of possible analysis states.
  - Therefore, ensemble perturbations ought to have the same statistics as analysis errors.
- Hypothesize that the differences between multi-center analyses represent realistic samples of analysis uncertainty.
- Let's compare ensemble perturbation statistics to multi-analysis differences.

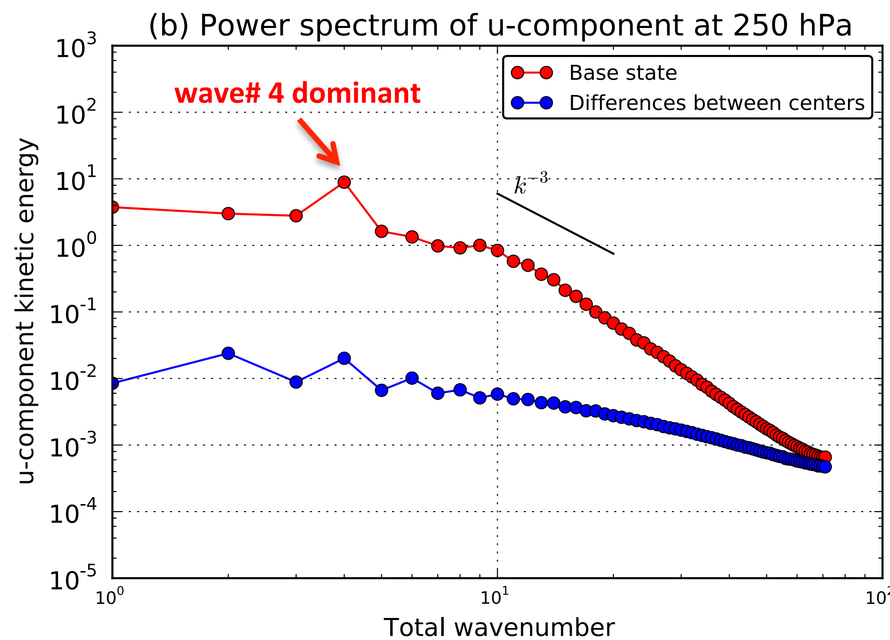
# Power spectra from analysis data

ECMWF used for base state; ECMWF - NCEP used for differences.
Spectra computed daily, then averaged over the full year.

T @ 500 hPa

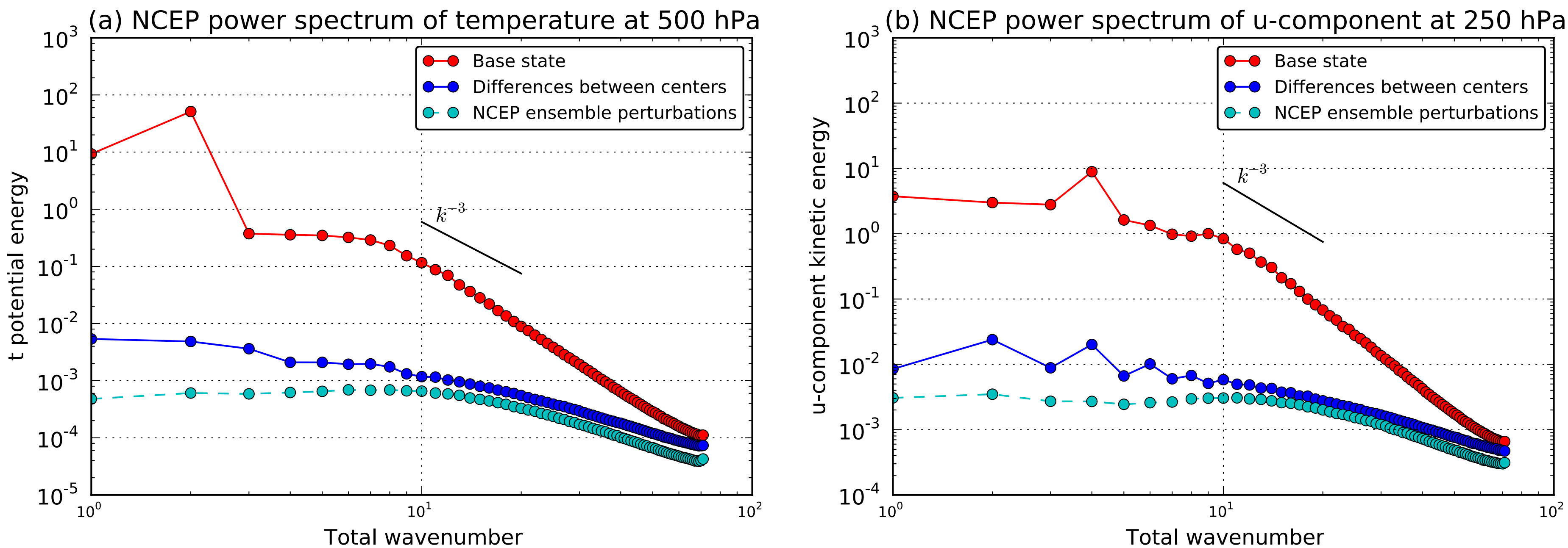



U @ 250 hPa




(1) Larger analysis "errors" (i.e., differences) at larger scales than at smaller scales, but …
(2) Large signal-to-noise ratio (S/N) at large scales, small S/N at smaller scales.
(3) Are analysis errors really that large at the largest scales? Probably yes for some models with larger biases, no for others with smaller biases (e.g., ECMWF).

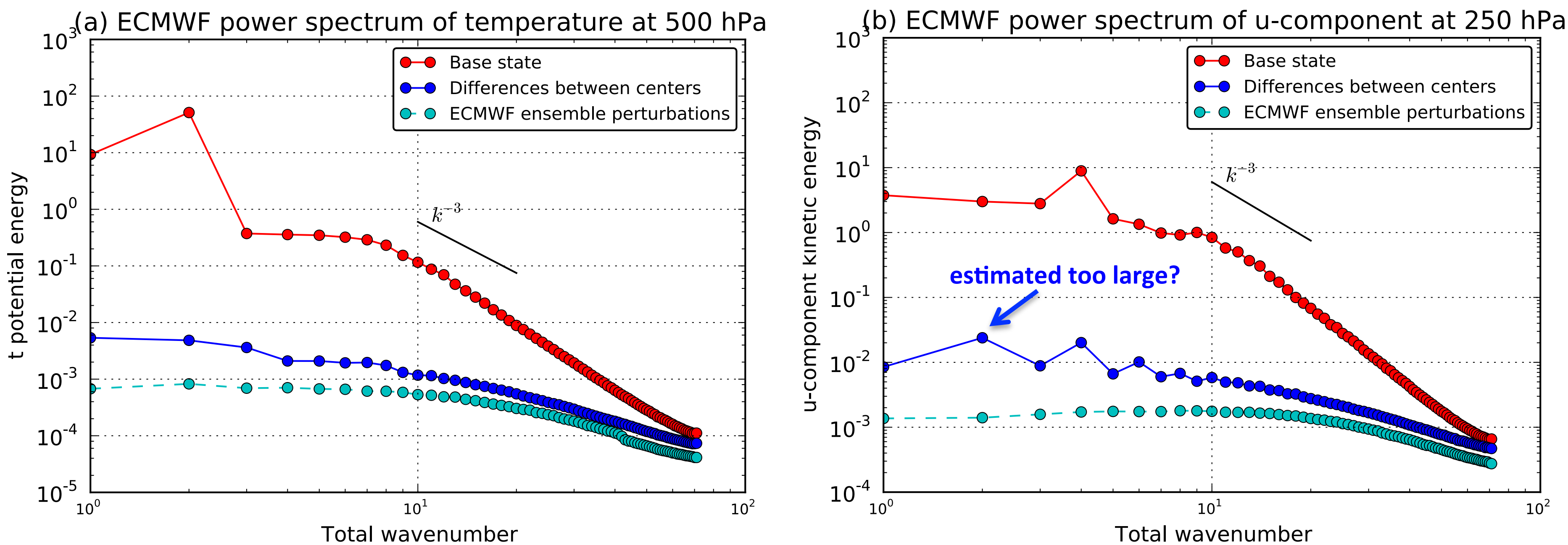
# Power spectra of ensemble perturbations, NCEP ensemble




(1) Suggests this ETR perturbation ensemble have insufficient power at planetary scales. This is consistent with the assumption made in the ETR that the analysis is unbiased while analyses between different centers suggest there is bias.

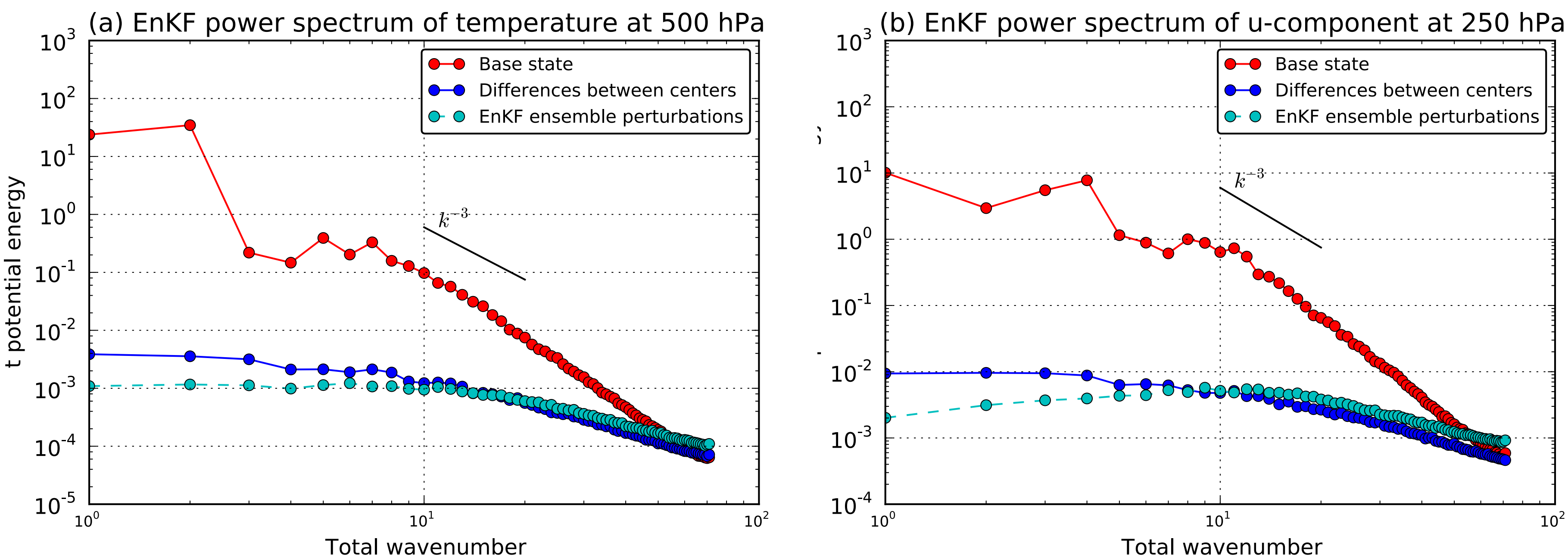
(2) ETR's underestimate of initial amplitude is the least for the small baroclinic scales. This may be because the breeding method inside the ETR generates perturbations that project onto the (finite amplitude) Lyapunov vectors, dominated by baroclinic scales.

# Power spectra of ensemble perturbations, ECMWF ensemble




(1) For wind, ECMWF's under-estimate of perturbation size is more dramatic than in NCEP's system. However, (a) their (energy norm) singular-vector perturbations may grow more rapidly, so this under-estimate may not be as pronounced in the forecast, and (b) various other diagnostics shown here today suggest ECMWF's analysis may have less model bias, hence the estimate of analysis error from difference between ECMWF and NCEP may overestimate the error spectrum for their system.

# Power spectra of ensemble perturbations, GFS EnKF

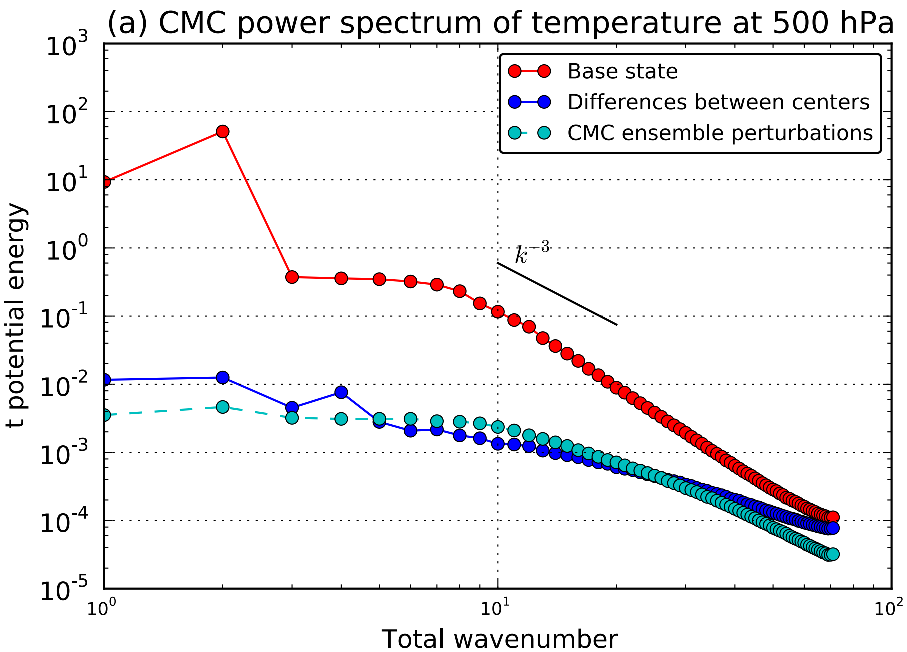


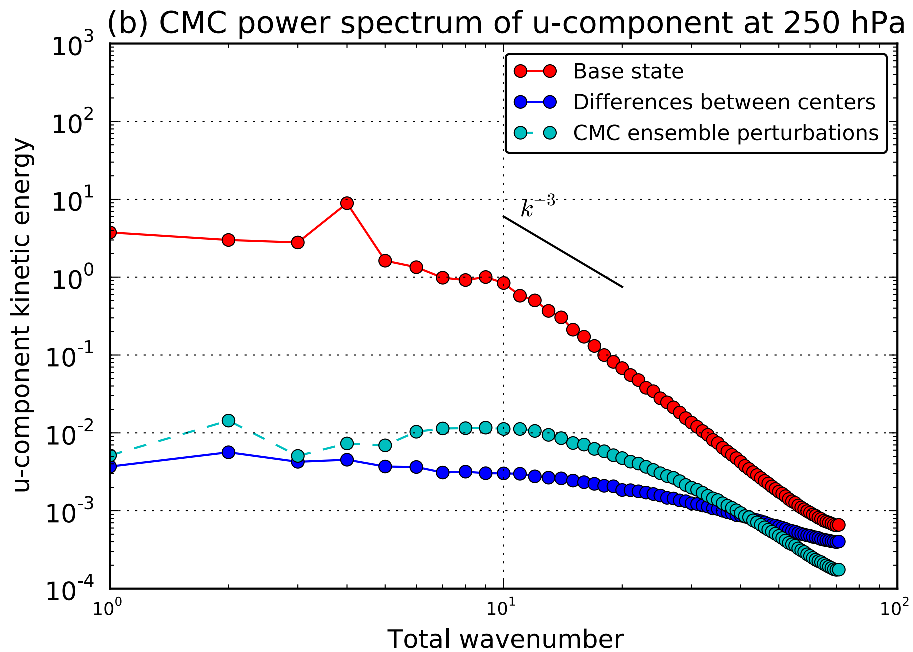

(1) Generally more power at all wavenumbers relative to ETR.

(2) Overestimate of power (i.e., amplitude of perturbations) at small scales. Likely this is attributable to inappropriate analysis increments due to the use of smaller-than ideal ensemble size (n=80) in the EnKF.

(3) Still some underestimate of power at large scales.

# Power spectra of ensemble perturbations, CMC ensemble



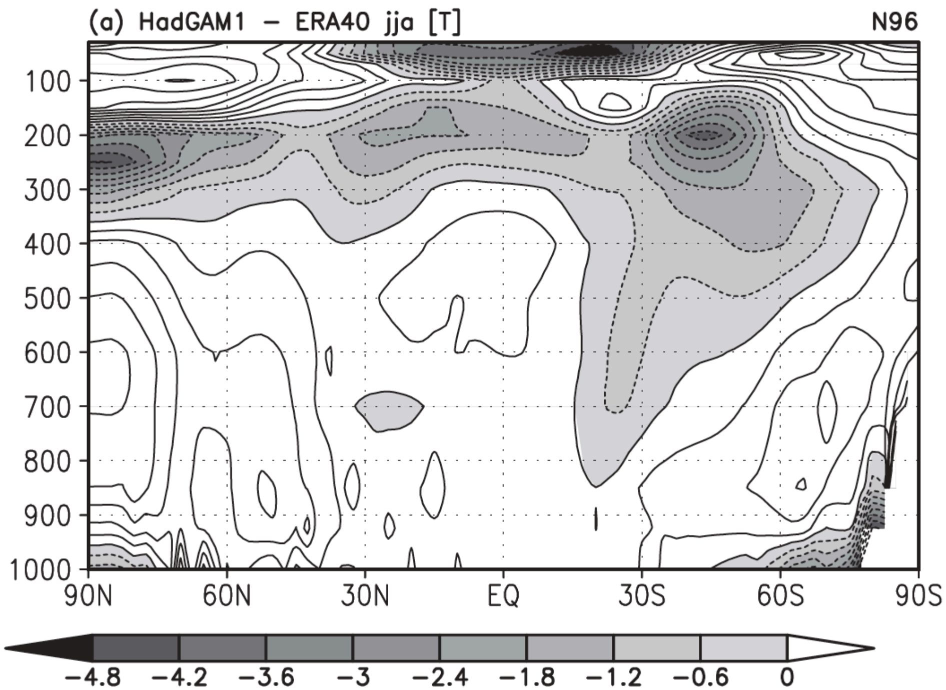

Reasonably well calibrated overall for t500; a bit of an overestimate of variance for u250.

# Conclusions

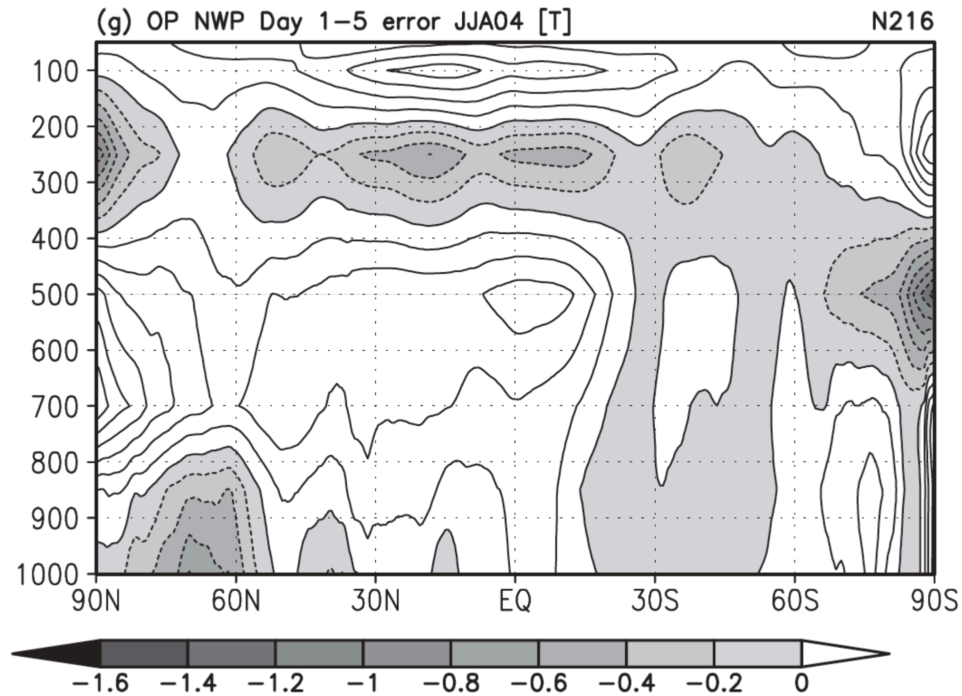
- Analyses, assumed to be unbiased, do exhibit substantial bias, as diagnosed from comparisons between analyses of different centers.
- Diagnostics suggest that NCEP's bias may be due in large part to issues related to data assimilation. There is greater consistency with other models after 24 h.
- Ensemble perturbations may be too small relative to multi-model estimates of analysis error, especially their size for the planetary scales. This under-estimate reflects the common assumption that the analysis is unbiased.
- Diagnostics such as these, leveraging TIGGE data, may prove useful for detecting forecast and assimilation system deficiencies. Consider such diagnostics for more widespread use?

# An aside: this sort of analysis underlies AMIP$^{T}$, where climate models are run in weather mode.

10-year climate forecast temperature bias with respect to reanalysis




JJA 2004 1-5 day forecast error, UKMO




model biases in climate models similar to what's observed in short-range forecast, so can use short-range forecast to improve and validate climate models.

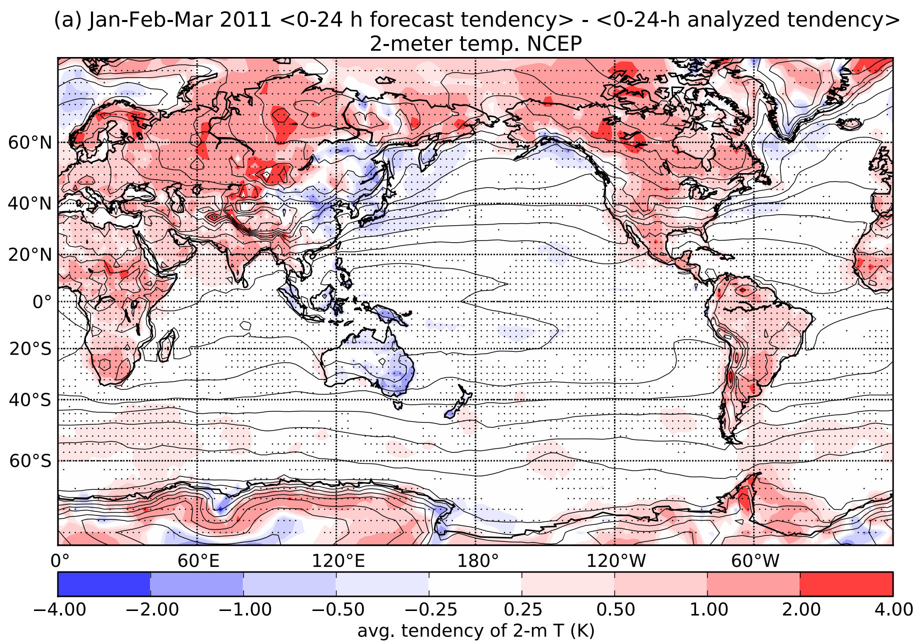


ref: Martin et al., *J. Clim*, 15 Nov 2010

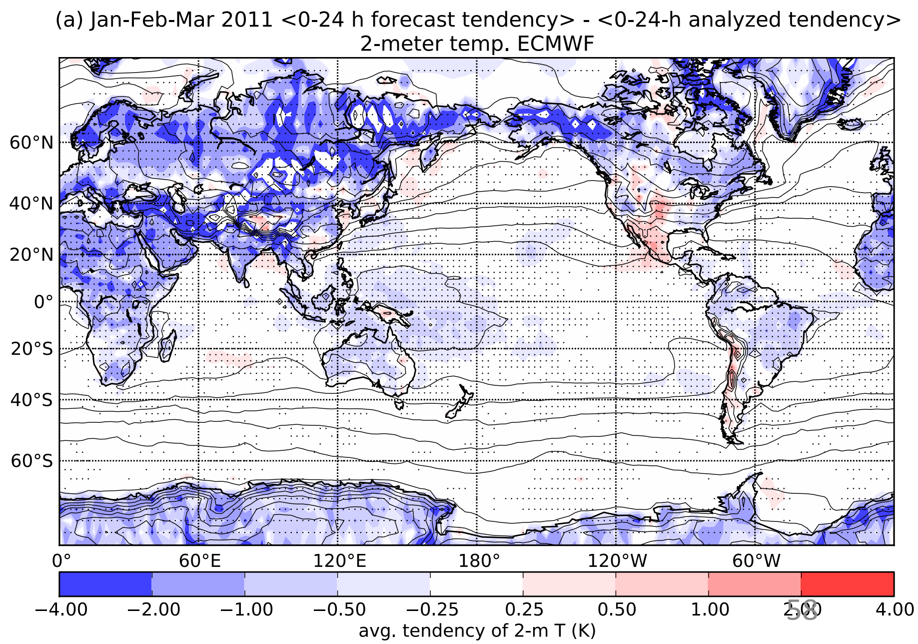
# Average 0-24 h tendencies in **2-m temp** <dF/dt> - <dA/dt>, Jan-Feb-Mar 2011

Dots convey mean is statistically significantly different than zero, via 2-sided *t*-test, $\alpha$=0.05 (field significance would be better).

NCEP

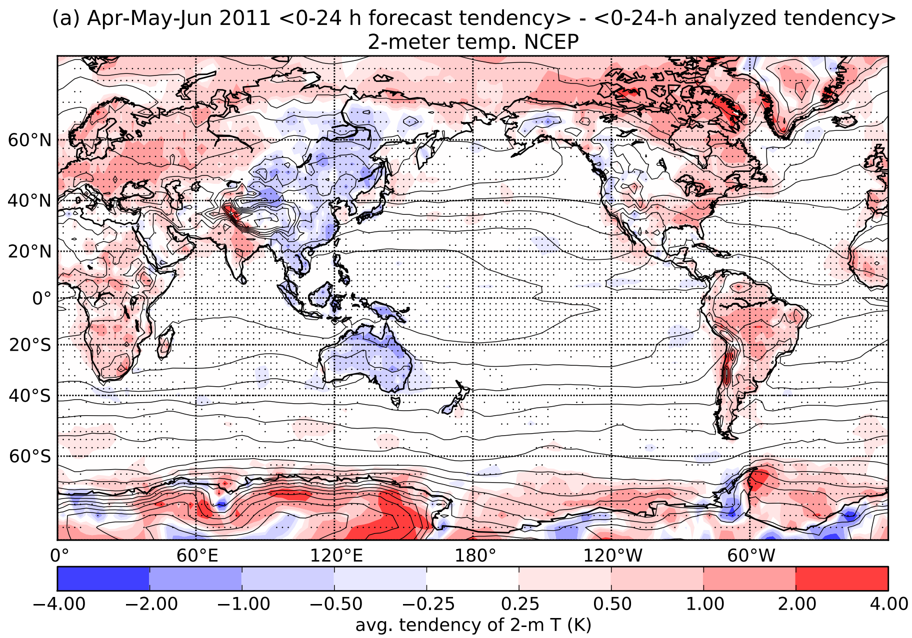



ECMWF

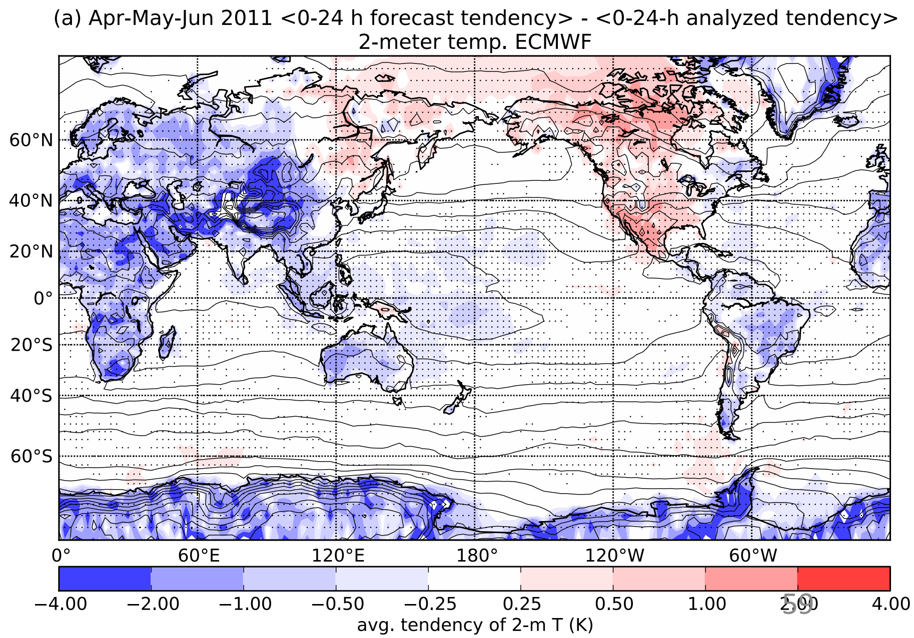


# Average 0-24 h tendencies in **2-m temp** <dF/dt> - <dA/dt>, Apr-May-Jun 2011

NCEP

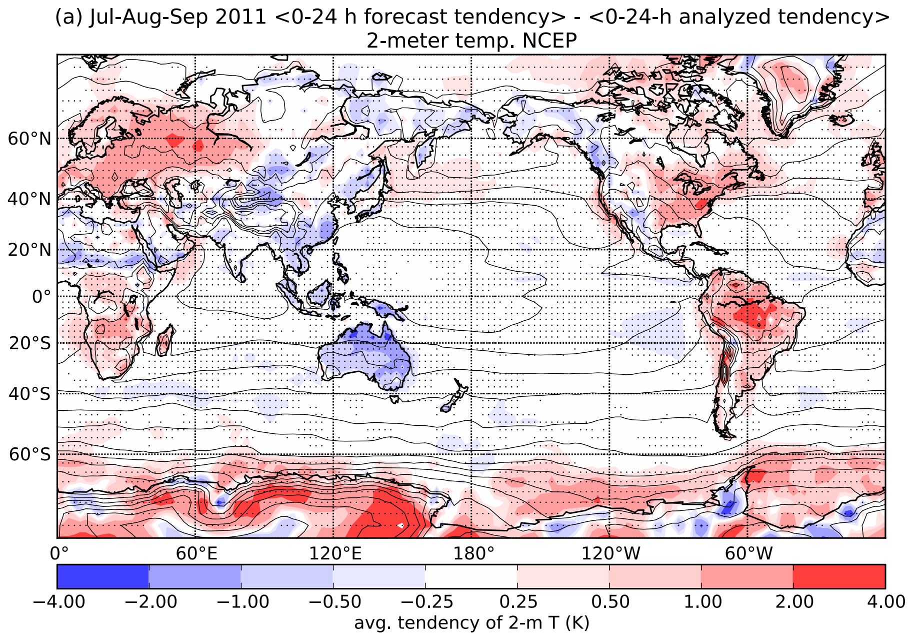



ECMWF

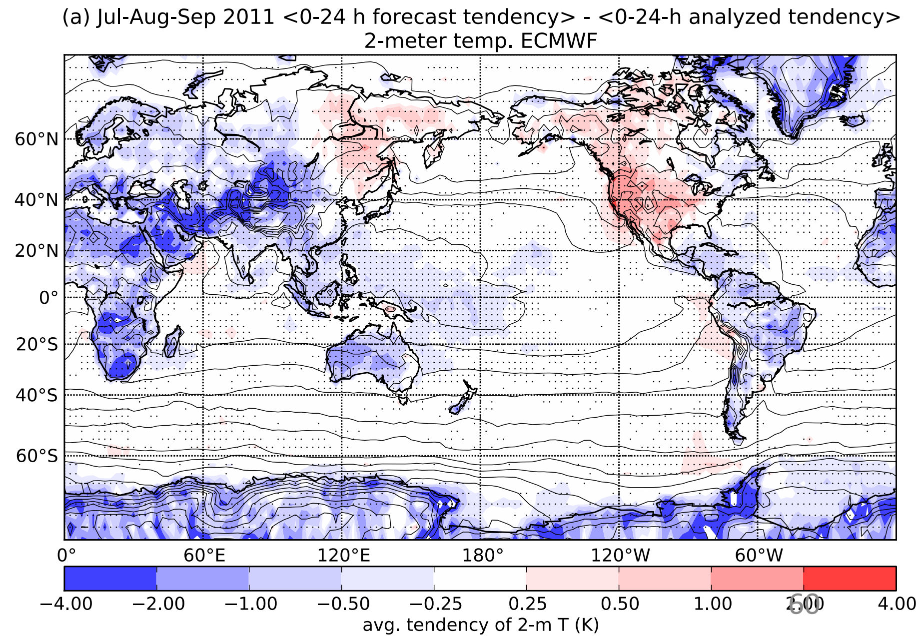


# Average 0-24 h tendencies in **2-m temp** <dF/dt> - <dA/dt>, Jul-Aug-Sep 2011

NCEP

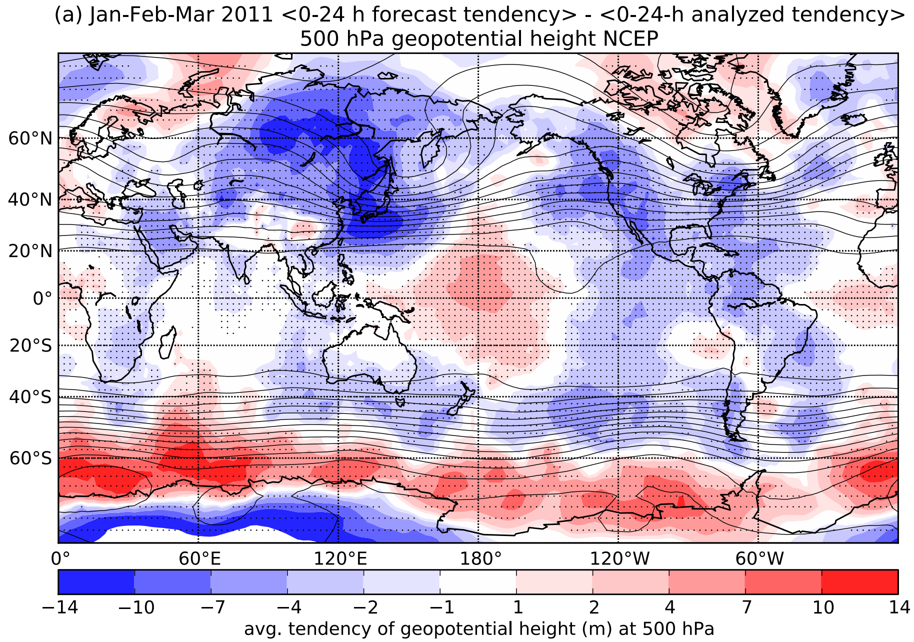



ECMWF

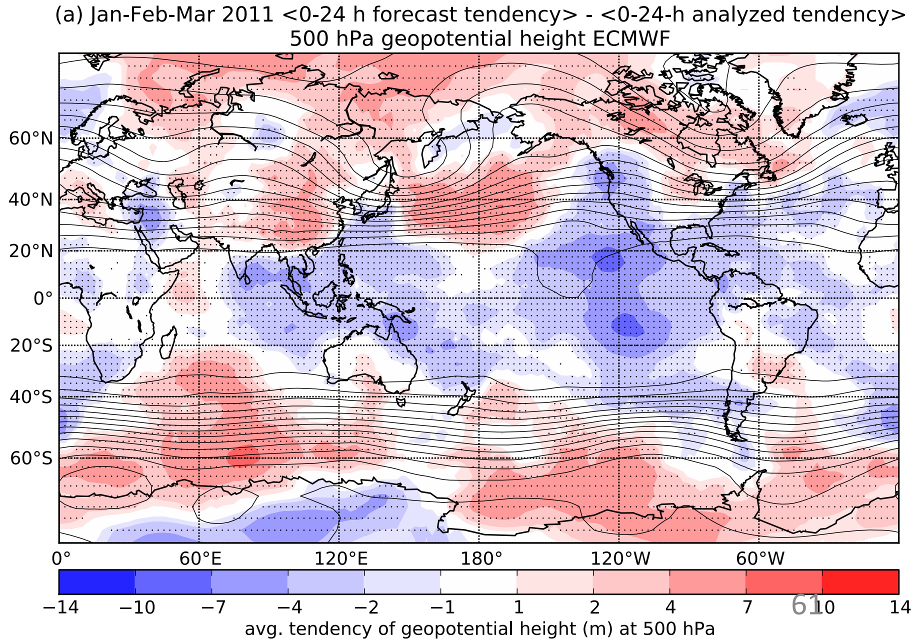


# Average 0-24 h tendencies in **500Z** <dF/dt> - <dA/dt>, Jan-Feb-Mar 2011

NCEP

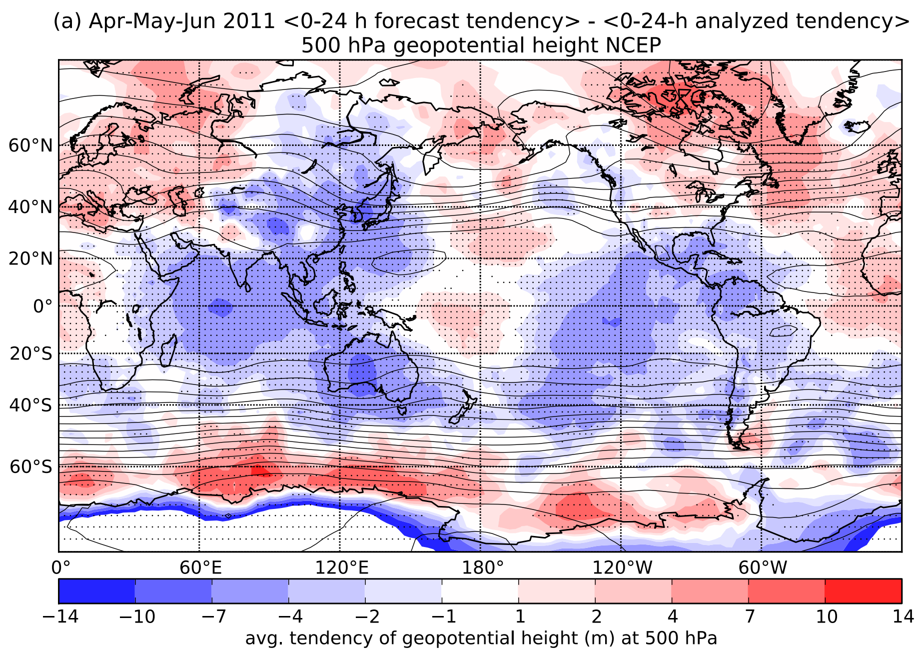



ECMWF

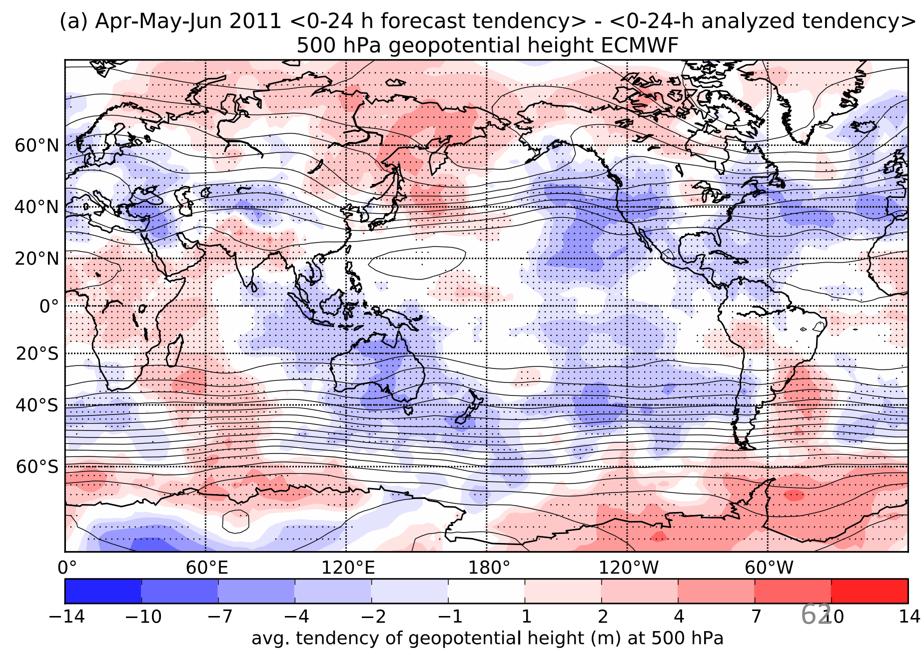


# Average 0-24 h tendencies in **500Z** <dF/dt> - <dA/dt>, Apr-May-Jun 2011

NCEP

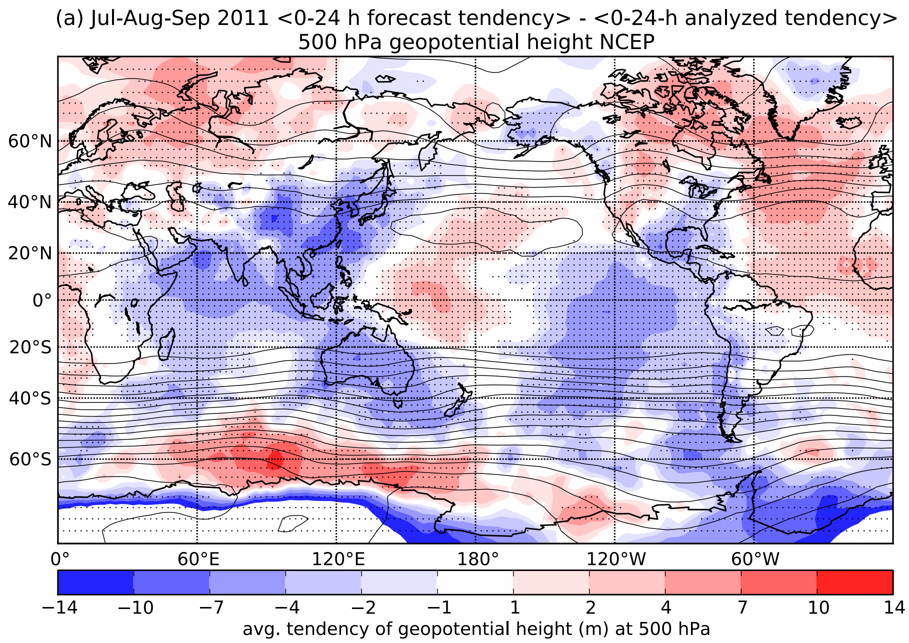



ECMWF

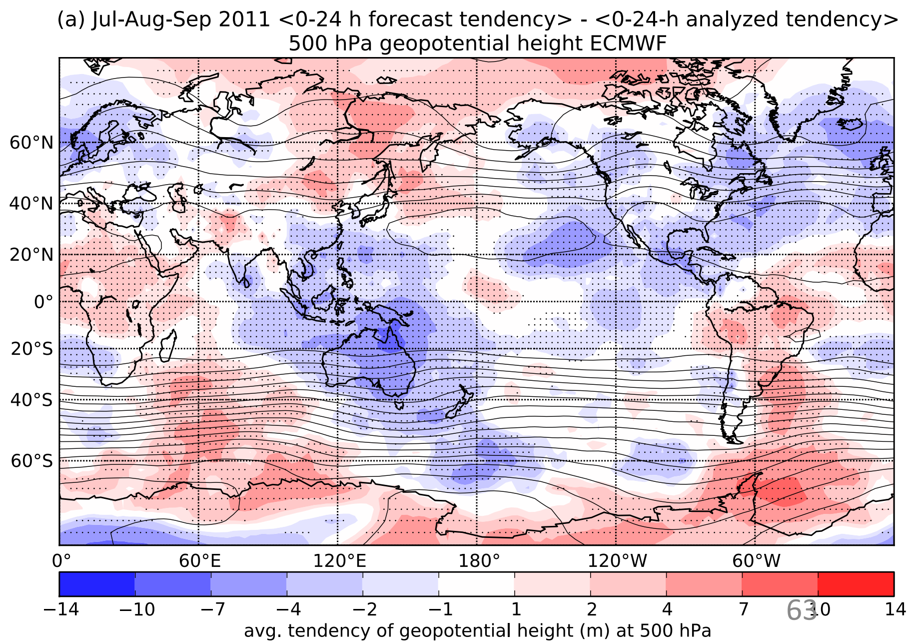


# Average 0-24 h tendencies in **500Z** <dF/dt> - <dA/dt>, Jul-Aug-Sep 2011

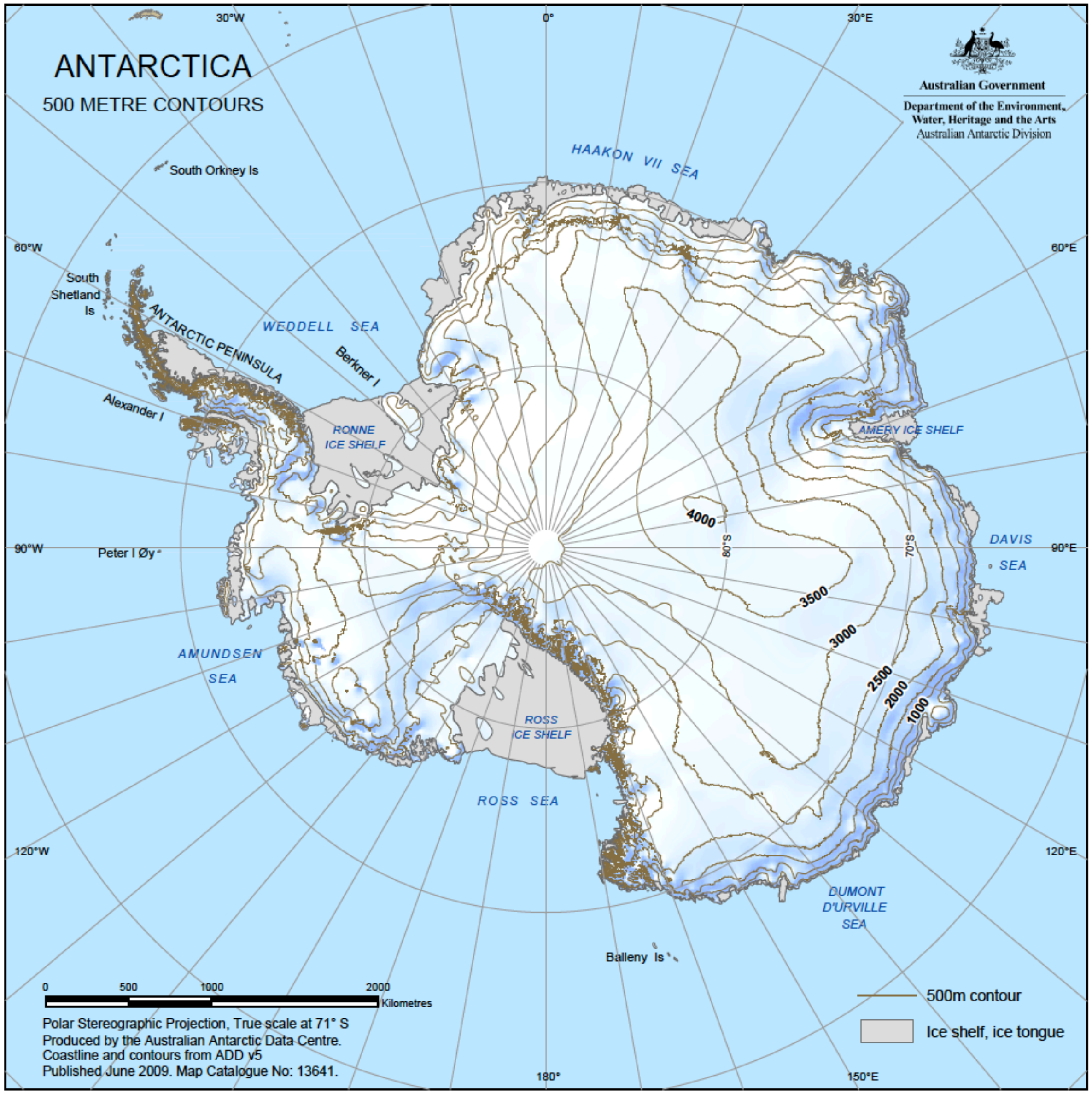
NCEP




ECMWF



# Recall: Antarctica's terrain



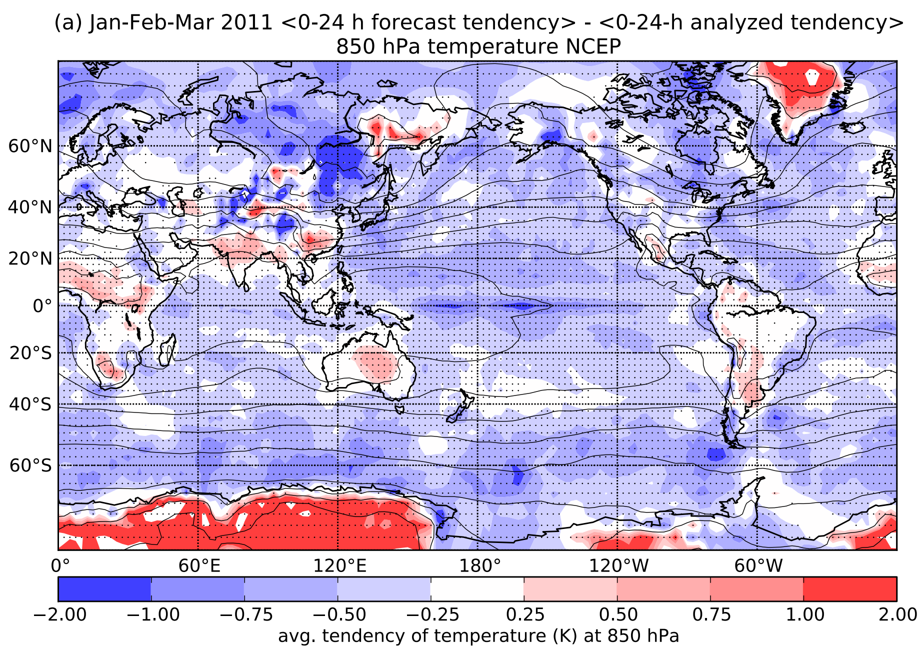

Typical 850 hPa geopotential height over Antarctica: ~ 1300 m. Hence nearly all of Antarctica $P_{sfc}$ < 850 hPa.

Typical 700 hPa geopotential height over Antarctica: ~ 2600 m. Approximately half of Antarctica $P_{sfc}$ < 700 hPa.

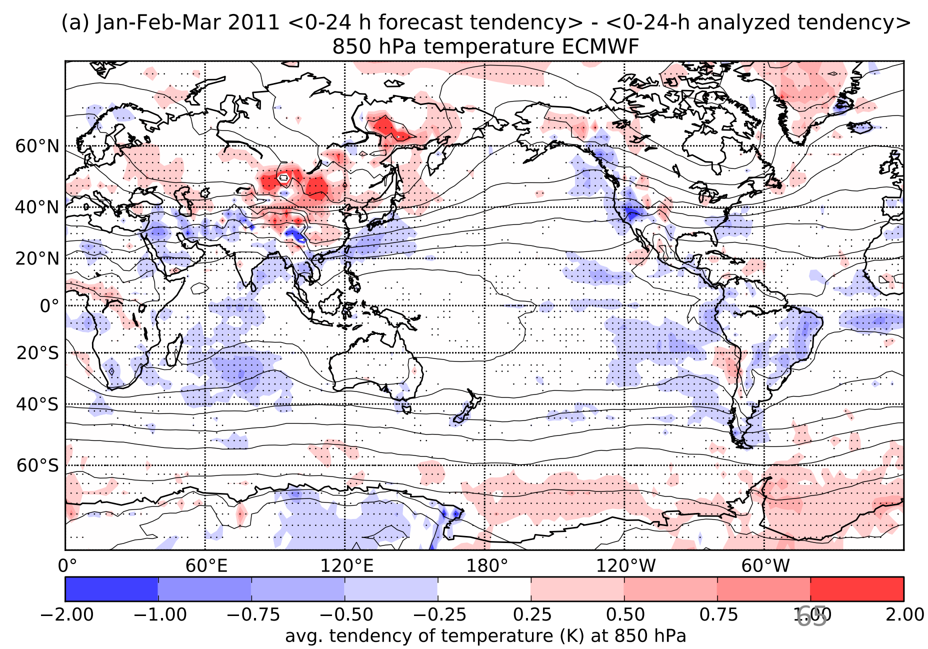
# Average 0-24 h tendencies in **850 T** <dF/dt> - <dA/dt>, Jan-Feb-Mar 2011

Negative NCEP tendencies over oceans happens for other seasons, too.

NCEP

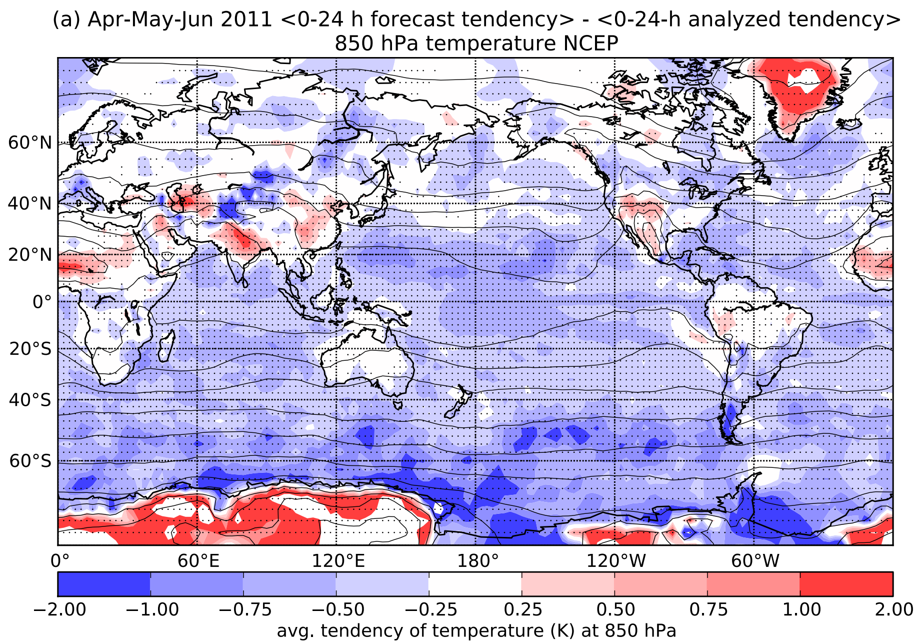



ECMWF

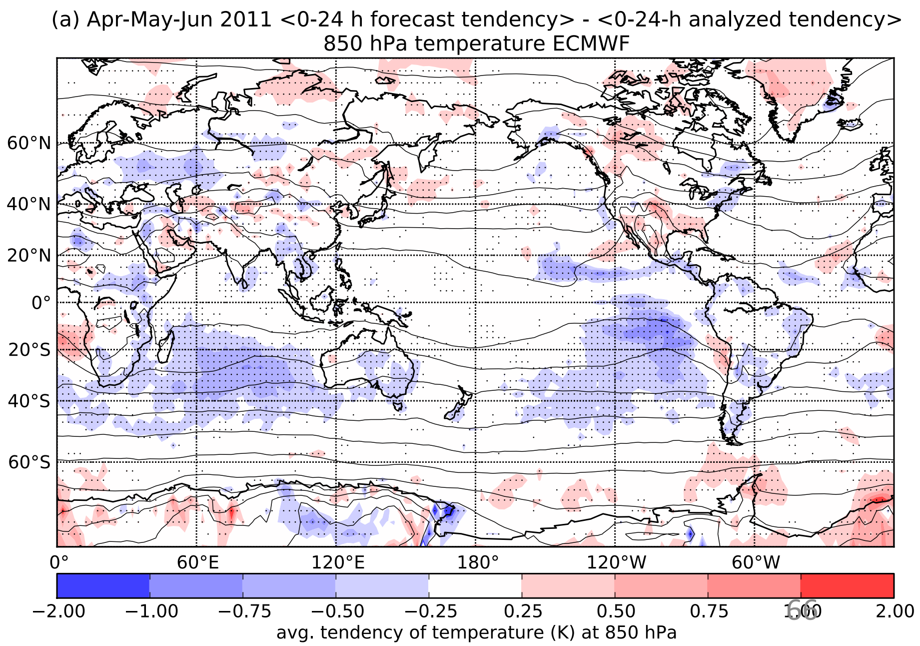


# Average 0-24 h tendencies in **850 T** <dF/dt> - <dA/dt>, Apr-May-Jun 2011

NCEP



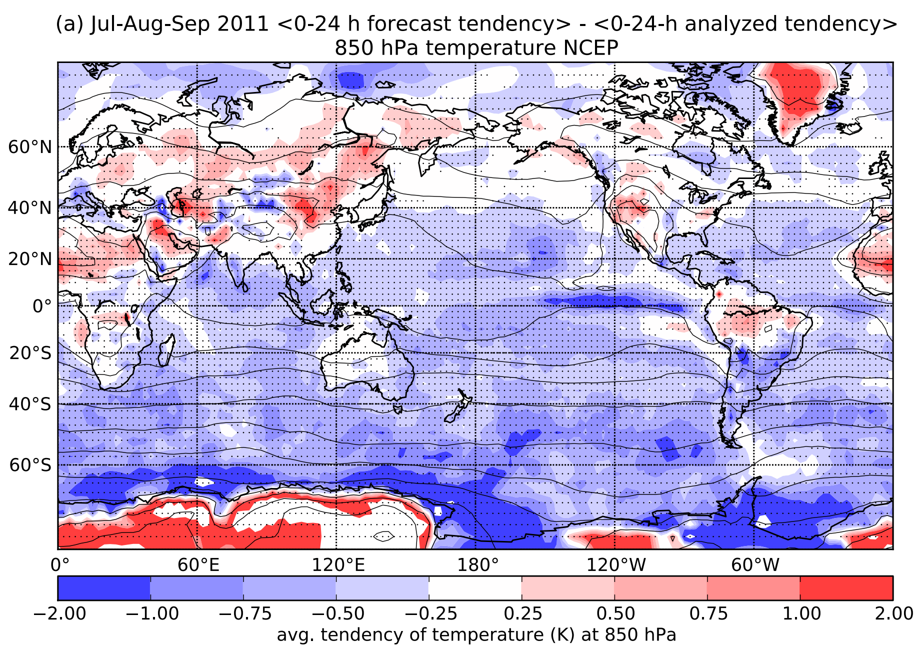

ECMWF



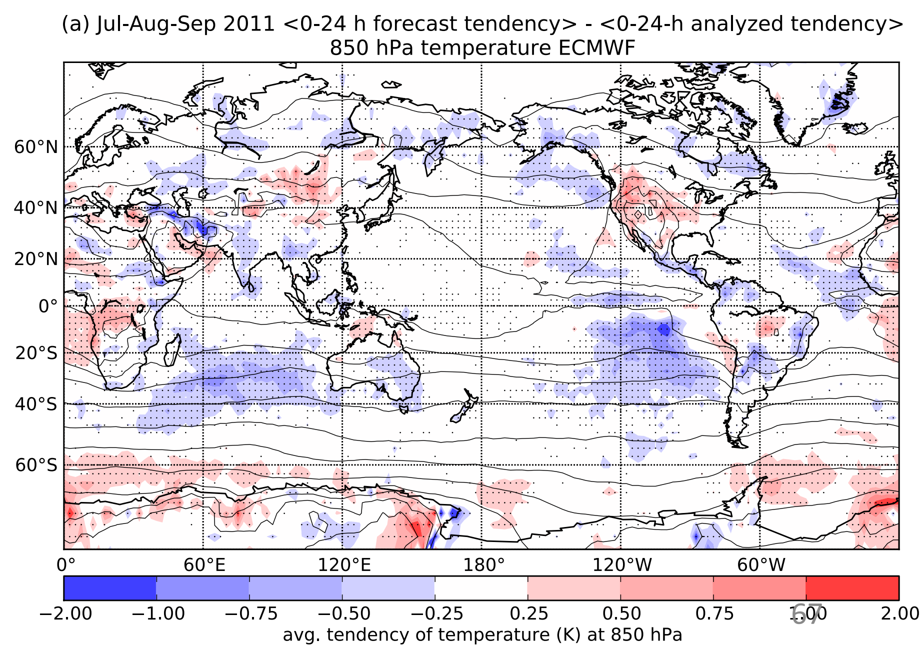
# Average 0-24 h tendencies in **850 T** <dF/dt> - <dA/dt>, Jul-Aug-Sep 2011

Are negative tendencies over the ocean for NCEP worse in the winter hemisphere?

NCEP



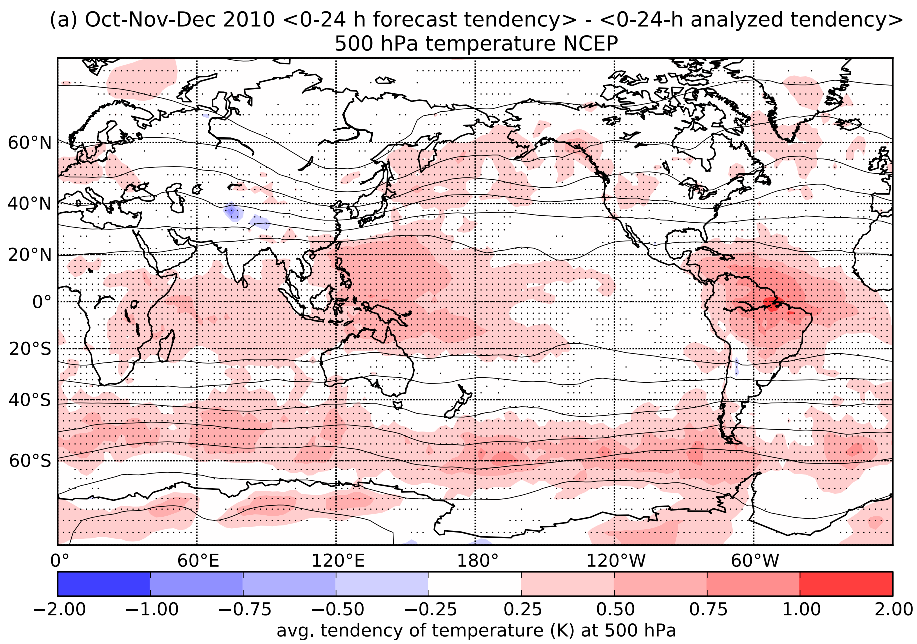

ECMWF



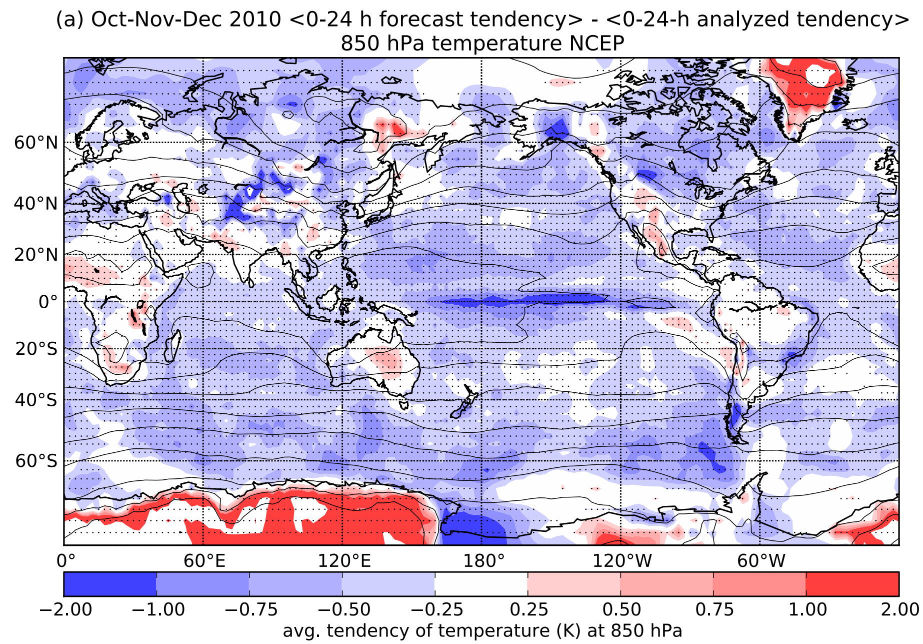
# Average 0-24 h tendencies in **850 & 500 T**
# <dF/dt> - <dA/dt>, Oct-Nov-Dec 2011

Implies a general stabilization of 850 to 500 hPa layer in short-term forecast.

NCEP T500




NCEP T850



# Recall: Nastrom & Gage (1984), energy spectra from plane observations

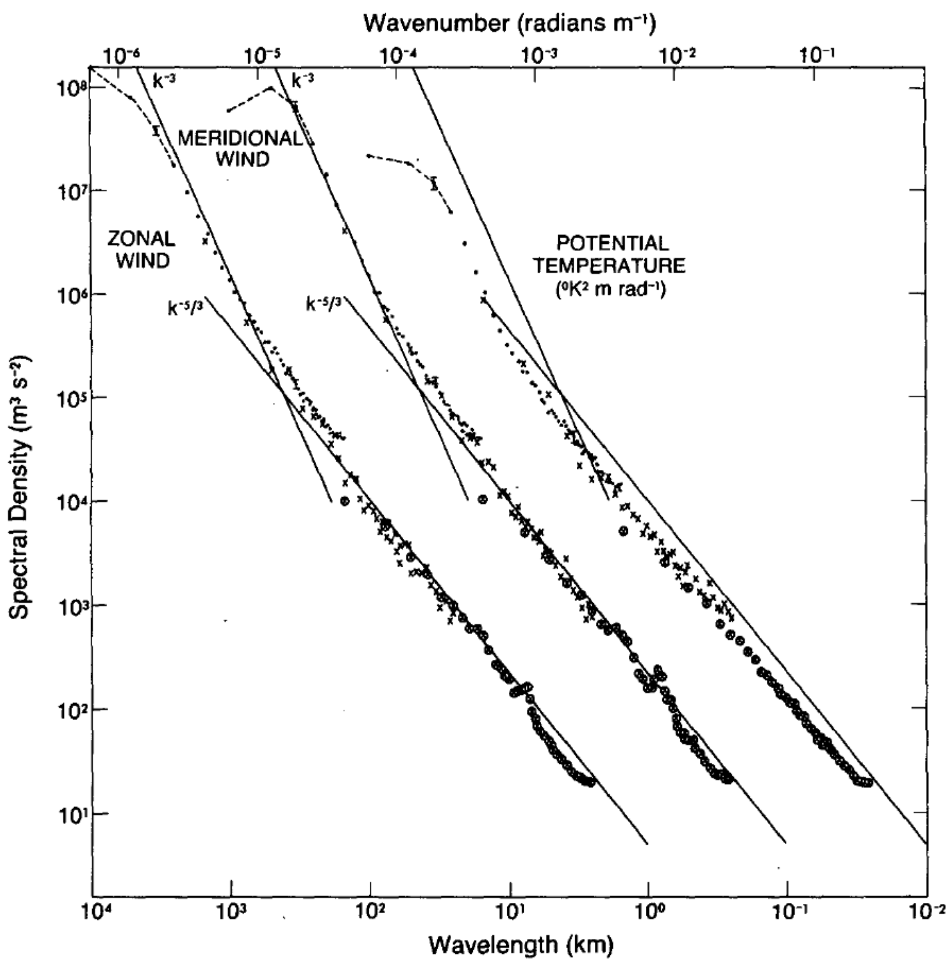



FIG. 3. Variance power spectra of wind and potential temperature near the tropopause from GASP aircraft data. The spectra for meridional wind and temperature are shifted one and two decades to the right, respectively; lines with slopes −3 and −5/3 are entered at the same relative coordinates for each variable for comparison.

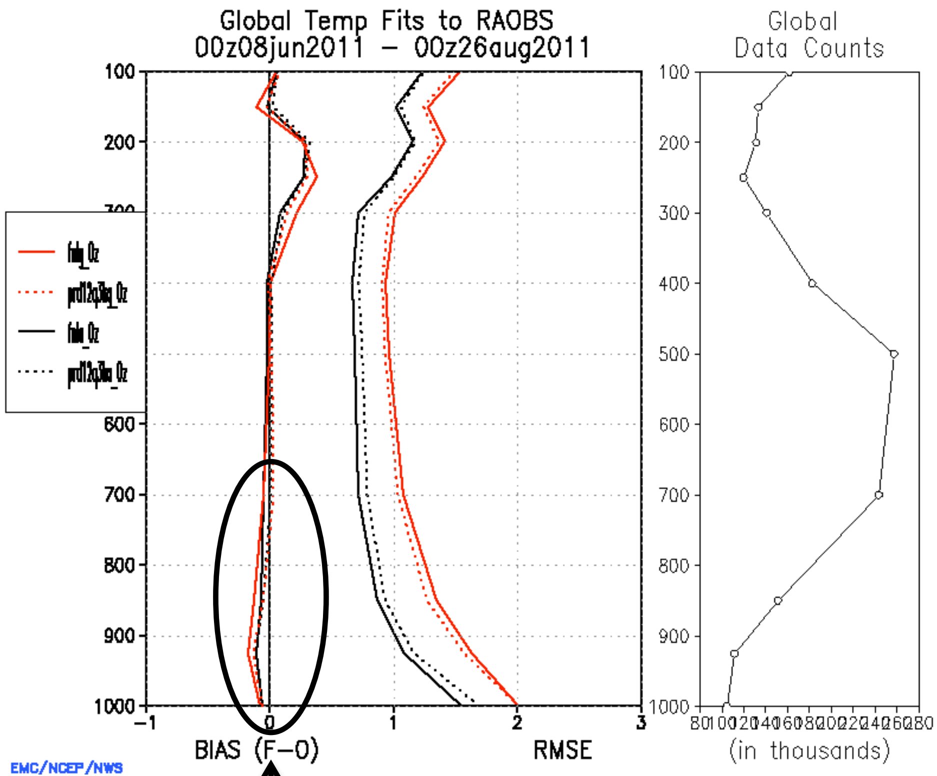
Now let's look at the power spectra.

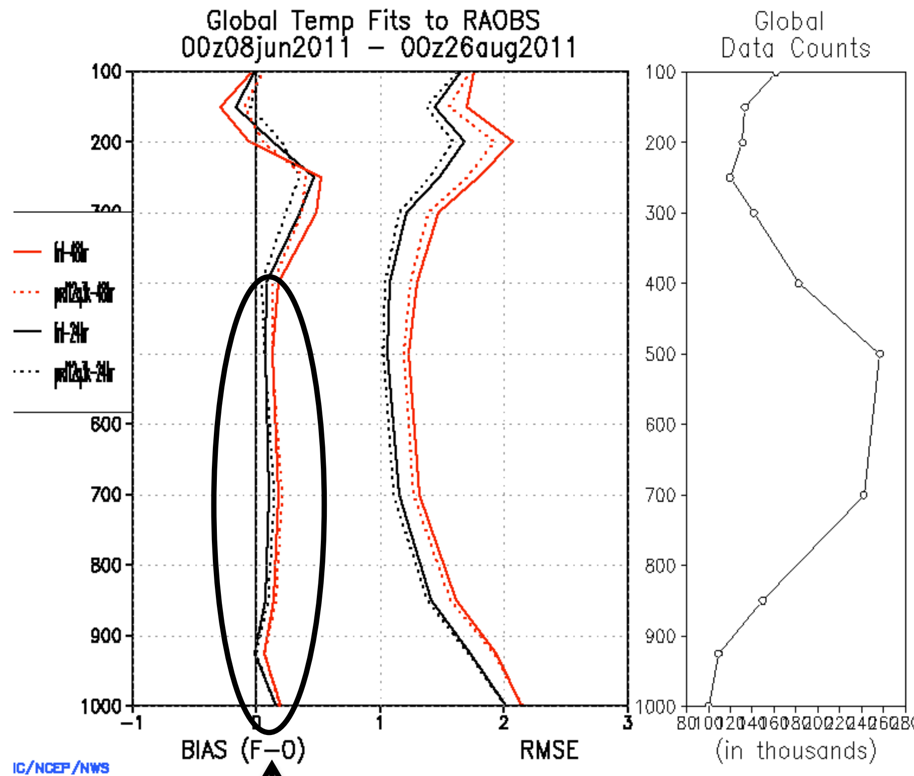
Recall that Nastrom and Gage experimentally confirmed the result that at the larger (synoptic) scales, the power spectra roughly follow a $k^{-3}$ power law, and at smaller scales, a $k^{-5/3}$ power law.

With the 2.5-degree analysis data used here, we can't expect to see much of the region of the spectra with $k^{-5/3}$.

# We'll end trying to understand what's behind this seemingly modest change

Recent statistics from parallel run that includes hybrid EnKF/GSI




At 0 hours, low-level temperature analyses are biased too cold.

At 24 h, temperature forecasts are biased too warm.